HUMBOLDT-UNIVERSITÄT ZU BERLIN

Experimental and theoretical analysis of X-chromosome inactivation as a paradigm for epigenetic memory and molecular decision-making

DISSERTATION

zur Erlangung des akademischen Grades

Doctor rerum naturalium
(Dr. rer. nat.)

eingereicht an der
Lebenswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von

Verena Mutzel

Berlin, 2021

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät:
Prof. Dr. Dr. Christian Ulrichs

Gutachter/innen:
1: Dr. Edda Schulz
2: Prof. Dr. Hanspeter Herzel
3: Dr. Luca Giorgetti

Tag der mündlichen Prüfung: 3.6.2021

Abstract

Different cell identities in different tissues of the same organism are made up of cells with identical genetic material but fulfill quite different functions. The identity of a cell is defined by the subset of genes that it expresses, which in turn is determined by the signals the cell has encountered in its past and is encountering in its present. Cells are thus able to *remember* information contained in past transient signals, and to convert this information into gene expression patterns. Therefore, multicellular life requires molecular memory that persists although the memory-initiating molecules are long gone. In some instances two copies of an identical gene within the same cell even acquire and maintain two different expression states, requiring local (in *cis*-encoded) memory. In this thesis, an interdisciplinary approach combining mathematical modeling and experiments is used to advance our understanding of how such *cis* memory can be generated. In the first (and major) part, the regulatory principles governing X-chromosome inactivation (XCI), a paradigm for epigenetic *cis* memory, are dissected. In the second part, mathematical modeling is used to investigate the potential of antisense transcription to generate such *cis* memory.

XCI is the mechanism for dosage compensation between the sexes in mammals. It is initiated through monoallelic upregulation of the long non-coding RNA Xist from one X chromosome, which mediates almost complete transcriptional silencing of this X chromosome. XCI regulation raises intriguing and thus far unanswered questions: How do cells count their X chromosomes and ensure that exactly one stays active? How do they make a mutually exclusive choice for one inactive X chromosome, and how do they then stably maintain this choice throughout subsequent cell divisions?

Using stochastic modeling, we show that XCI onset only requires two regulators: A *trans*-acting Xist activator that ensures female specificity and a *cis*-acting Xist repressor that allows stable maintenance of alternative Xist expression states. This two-regulator network can recapitulate Xist expression patterns across different species and makes a novel prediction that is validated experimentally: Cells are able to revert biallelic Xist expression to monoallelic expression. With a mechanistic stochastic model we show that Xist's antisense transcript Tsix might be the *cis*-acting Xist repressor, uncovering the molecular mechanism behind the stabilization of the alternative Xist expression states. Building upon Tsix' possible functional role in stabilizing alternative Xist expression states on the active and inactive X chromosome, the second part of this thesis investigates the potential of antisense transcription to maintain a transient transcriptional memory. We find that mutual repression between a pair of antisense genes can allow the locus to remember the transcription state it has acquired due to a past signal for several days.

Zusammenfassung

Vielzelliges Leben zeichnet sich durch Aufgabenteilung aus. Zellen mit identischer genetischer Information bilden alle unterschiedlichen Gewebe eines Organismus. Ihre Identität ist bestimmt durch die Gene, die sie exprimieren. Welche Gene exprimiert werden hängt wiederrum von den Signalen ab, die eine Zelle in ihrer Vergangenheit erhalten hat. Deshalb braucht jedes vielzellige Leben, in dem unterschiedliche Zelltypen unterschiedliche Aufgaben erledigen, ein Gedächtnis, das jeder Zelle erlaubt, sich an ihre Identität und damit an ihre Aufgaben zu erinnern. Eine Leberzelle sollte zum Beispiel möglichst nicht vergessen, dass sie eine Leber- und keine Nierenzelle ist. In Extremfällen, in denen es nötig ist, dass zwei Kopien desselben Gens unterschiedlich exprimiert werden, braucht die Zelle sogar ein lokal (in *cis*) kodiertes Gedächtnis, das jeder Genkopie erlaubt sich an ihren Expressionszustand zu erinnern.

Diese Arbeit verwendet einen interdisziplinären Ansatz, um zu verstehen, wie solch ein *cis*-Gedächnis entstehen kann. Der Hauptteil der Arbeit befasst sich mit der Regulation der X-Chromosom-Inaktivierung (XCI), einem Paradigma für solch *cis*-kodiertes Gedächtnis. Inspiriert durch die Resultate dieser Analyse untersuchen wir im zweiten Teil, ob und unter welchen Bedingungen Antisense-Transkription ein *cis*-Gedächtnis generieren kann.

XCI ist der Mechanismus, den Säuger zur Dosiskompensierung zwischen weiblichen und männlichen Zellen verwenden. XCI wird ausgelöst durch die monoallelische Hochregulation der langen nicht-kodierenden RNA Xist von einem der zwei X-Chromosomen in weiblichen Zellen. Die Xist RNA vermittelt dann das Ausschalten der Gene auf diesem X-Chromosom. Das wirft einige interessante Fragen auf: Wie zählen Zellen ihre X-Chromosomen und stellen sicher, dass genau eines aktiv bleibt? Wie entscheiden sie, welches X-Chromosom aktiv bleibt und welches ausgeschaltet wird? Und wie erinnern sie sich an diese Entscheidung und behalten sie stabil bei durch alle weiteren Zellteilungen?

Mithilfe eines stochastischen Modells zeigen wir, dass diese XCI Regulation prinzipiell durch nur zwei Regulatoren erklärt werden kann: Ein global (in *trans*) agierender XCI Aktivator und ein lokal (in *cis*) agierender XCI Repressor. Dieses Netzwerk aus nur zwei Regulatoren kann die Xist Expressionsmuster in verschiedenen Säugerspezies reproduzieren, von der Maus bis zum Mensch. Es sagt außerdem voraus, dass Zellen in der Lage sind, biallelische zu monoallelischer Xist Expression zu korrigieren, eine Vorhersage, für die wir tatsächlich experimentelle Belege finden. Mit einem mechanistischen Modell zeigen wir, dass das *cis*-Gedächtnis über den Xist Expressionszustand durch Antisense-Transkription zustande kommen könnte. Auf dieser Hypothese aufbauend untersucht der zweite Teil der Arbeit das Potential von Antisense-Transkription, ein

lokales Gedächtnis über den Expressionszustand eines Gens zu generieren, genauer. Diese Analyse sagt vorher, dass Antisense-Repression den Expressionszustand eines Lokus tatsächlich für einige Tage stabil erhalten kann.

Abbreviations

aXR	Autosomally encoded Xist repressor
cXR	*cis*-acting Xist repressor
Edu	5-ethynyl-2'-deoxyuridine
Fsw	First-switching-time
HDR	Homology directed repair
KO	Knock-out
mESCs	Mouse embryonic stem cells
MFS	Minimal first-switching-time
Mut	Mutant
Oc	Occlusion
PC	Polymerase collisions
PolII	RNA polymerase II
PP	Pluripotency factors
PR	Promoter repression
PRC2	Polycomb repressive complex 2
RNAP	RNA polymerase
SDC	Sitting duck complex
SDI	Sitting duck interference
SNP	Single-nucleotide polymorphisms
ssODN	Single-stranded oligo deoxynucleotide
TAD	Topologically associating domain
TF	Transcription factor
TI	Transcriptional interference
TSS	Transcription start site
tXA	*trans*-acting Xist activator
WT	Wild type
Xa	Active X chromosome
XaXa	State of a female cell with two active X chromosomes
XaXi	State of a female cell with one active and one inactive X chromosome
XCI	X-chromosome inactivation
Xi	Inactive X chromosome
Xic	X-inactivation center
Xm	Maternal X chromosome
XO	Female genotype with only one X and no Y chromosome
Xp	Paternal X chromosome

Contents

Part I

A regulatory model for X-chromosome inactivation

1 Introduction

"The present communication suggests that the evidence of mouse genetics indicates: (1) that the heteropyknotic X-chromosome can be either paternal or maternal in origin, in different cells of the same animal; (2) that it is genetically inactivated." [1]

These are the opening lines of the concise one page summary, in which Mary F. Lyon for the first time proposed the process of random X-chromosome inactivation, or what was to become the Lyon hypothesis. What Lyon had discovered was the mammalian solution for achieving dosage compensation between the sexes. Every species that has evolved genetic sex determination by distinct sex chromosomes faces the need for compensating the dosage imbalance between the sexes. While the mammalian Y is a small chromosome with few genes, many of which are involved in male sex determination and development, the mammalian X is a much bigger chromosome. It contains around 1000 genes that mostly fulfill non-sex-specific functions, but that are still present as two copies in females and only as a single copy in males. Different species have evolved alternative strategies to tackle this problem: In the fruit fly *Drosophila melanogaster*, gene expression from the single X in males is upregulated [2, 3] while the nematode *Caenorhabditis elegans* downregulates expression from both X chromosomes in females [4, 5]. Mammals have evolved yet another and possibly even more astonishing mechanism of dosage compensation: X-chromosome inactivation (XCI), which results in the (almost) complete transcriptional inactivation of one of the two X chromosomes in females. In eutherian mammals, the choice of the inactive X chromosome is even random (independent of the parental origin), making females mosaics as already observed by Lyon, with profound implications for X-linked diseases [1]. Major insight into the regulation of XCI has come from the use of female mouse embryonic stem cells (mESCs), which are derived from the inner cell mass of the blastocyst, possess two active X chromosomes in the naive state, and recapitulate random XCI upon differentiation.

1.1 The X-inactivation center

The regulatory locus controlling XCI is the *X-inactivation center (Xic)*, a genomic region on the X chromosome that has been defined as the minimal region sufficient to trigger XCI if present in two copies (Fig 1.1) [6–8]. At the heart of the *Xic* lies the *Xist* locus, which encodes the master regulator of XCI: the long non-coding RNA Xist. Xist is upregulated from the future inactive X chromosome (Xi) and gradually coats the complete X chromosome in *cis*, mediating its transcriptional shut-down. The Xist RNA is a very long transcript (17 kb), whose sequence is poorly conserved between species except for a number of repeat rich regions termed

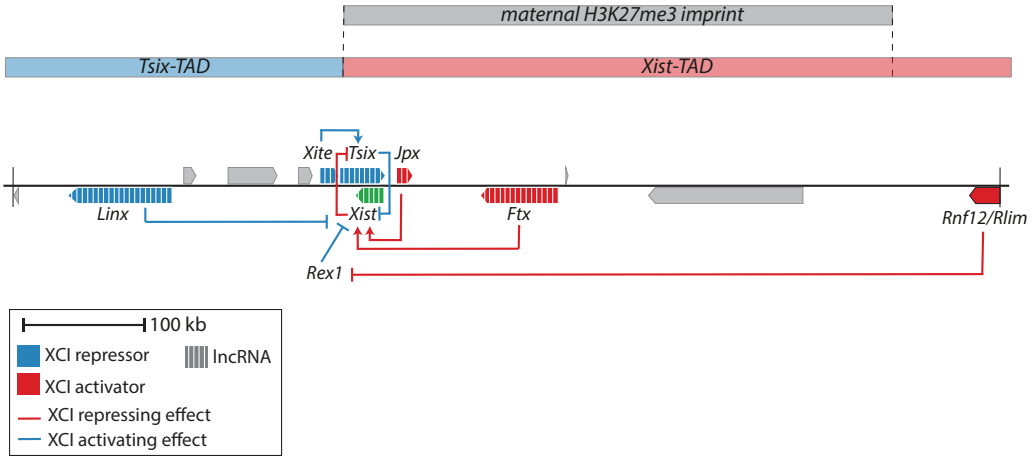

FIGURE 1.1: The mouse *X-inactivation center (Xic)*. Schematic representation of the region chrX:103184059-103981285 (mm10) in the mouse genome. Genes colored in red and blue indicate Xist activators and repressors, respectively. A striped pattern marks lncRNA genes. Activating and inhibitory arrows indicate known interactions between XCI regulators. Red links activate Xist directly or indirectly by repressing its repressors, blue links repress Xist directly or indirectly by activating its repressors.

repeats A-F [9–11]. The A repeat in the 5' part of Xist shows the strongest conservation and mediates silencing through the recruitment of transcriptional repressors such as Spen (see section 1.4) [12]. Other repeat regions are essential for Xist RNA association with chromatin, or to recruit components of other silencing pathways such as PRC2 (reviewed in [13]). The Xist RNA-mediated transcriptional shut-down of one X chromosome is accompanied by a series of remarkable changes such as the refolding of the entire chromosome [14–18] and the formation of a repressive compartment devoid of RNA polymerase II (PolII) and transcription factors (TFs) (see section 1.4).

1.2 Xist regulators

Apart from *Xist* the *Xic* harbors several coding and non-coding loci that are involved in Xist regulation. Curiously, their functional role in XCI is reflected in their spatial organization: The *Xic* is divided into two domains of increased self-interaction, one domain of positive and one of negative Xist regulators [19]. The *Xist* gene is located at the boundary between these two topologically associating domains (TADs) (Fig 1.1).

The best studied negative regulator of Xist is its antisense gene *Tsix*, that also encodes a long non-coding RNA (lncRNA). *Tsix* spans the complete *Xist* gene, transcribing beyond the *Xist* promoter (Fig 1.1). Tsix clearly functions as a *cis* repressor of Xist, as heterozygous *Tsix* deletions skew the choice of the inactive X towards the mutant allele [20–23], likely because they accelerate Xist upregulation in *cis*. Homozygous *Tsix* mutations have been reported to result in an increased fraction of biallelically Xist expressing cells [24]. Mechanistically, Tsix transcription establishes a repressive chromatin environment at the Xist promoter [25–28]. In post-XCI cells, Xist and Tsix show opposing expression patterns on the active (Xa) and the inactive (Xi) X chromosome, with Tsix being expressed almost exclusively from the Xa and Xist being expressed from the Xi. It is unclear whether Tsix is conserved in other species. Apart from mice, Tsix

transcription so far has only been detected in human embryonic carcinoma cells, but with a reduced overlap between XIST and TSIX, such that TSIX does not transcribe through the XIST promoter [29].

Interestingly, in humans another lncRNA termed XACT antagonizes XIST in *cis* and its expression pattern in primed hESCs is reminiscent of that of Tsix in mouse: XACT coats the Xa and XIST coats the Xi [30, 31]. Importantly, however, XACT is not localized antisense to XIST but lies several million bp away. Two other lncRNA loci within the *Xic* have also been reported to act as *cis* repressors: Xite acts as a Tsix enhancer (Fig 1.1). Its deletion also results in mildly skewed Xi choice, likely an indirect effect that is caused by Tsix downregulation in *cis* [32, 33]. Linx functions as a *cis*-regulatory DNA element that represses Xist independently of its transcription (Fig 1.1) [34].

Among the positive Xist regulators, one of the best characterized is the E3 ubiquitin ligase Rnf12 which activates Xist by targeting the pluripotency factor and Xist repressor Rex1 for degradation (Fig 1.1) [35–37]. Rnf12 is also encoded within the *Xic*, and is therefore expressed at higher levels in females than in males. It has been shown to activate Xist in a dose-dependent manner and thereby contributes to ensuring female-specific XCI. Its overexpression in males indeed results in ectopic XCI, while its heterozygous deletion in females does not abolish but only delay XCI [35]. Complete absence of Rnf12 prevents imprinted XCI in mouse embryos, but the random wave was reported to be unaffected [38–40]. In mESCs the Rnf12 knock-out (KO) phenotype seems to depend on the precise culture conditions: Xist upregulation fails in some but not all conditions [38, 39, 41].

Another XCI activator encoded in the *Xic* is the lncRNA Jpx (Fig 1.1). Whether Jpx acts in *cis* or *trans* still remains controversial. Jpx RNA has been proposed to activate Xist by evicting the Xist repressor Ctcf from the *Xist* locus via molecular titration [42]. However, it is difficult to imagine how a rather weakly expressed RNA could titrate away a highly expressed protein, or how this titration could be limited to the *Xist* locus while other loci would still be bound by Ctcf. In addition, other studies have found Jpx transgenes to be unable to compensate for the deletion of the endogeneous locus, arguing against an action in *trans* [43]. Jpx's mechanism of action also seems to have diverged in mouse and human: In mouse the Jpx transcript is essential, while in humans Jpx transcription activates Xist in *cis* [44]. Jpx is one of the few genes that escapes Xist-mediated silencing. Ftx is another lncRNA in the *Xic* that activates Xist in *cis* by the act of transcription (Fig 1.1) [45].

Several XCI regulators have thus been identified over the years and for some their precise mechanism of action is known. It however still remains elusive how these regulators act together to ensure correct XCI onset.

1.3 What times XCI?

The onset of XCI during embryonic development is precisely timed. It occurs during the transition from naive to primed pluripotency, a time period which is accompanied by major global

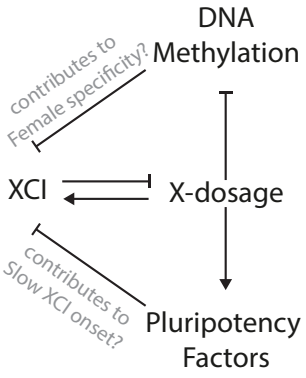

FIGURE 1.2: Global effects of X-chromosome dosage. Double X-dosage induces XCI, but also maintains pluripotency factor expression and delays global DNA methylation, thereby delaying differentiation in females. In males, faster DNA methylation could contribute to Xist repression, while in females slower downregulation of pluripotency factors may reduce the probability for Xist upregulation, thereby ensuring proper XCI timing and robust monoallelic XCI.

epigenetic and transcriptional changes, such as the downregulation of naive pluripotency factors, the upregulation of primed pluripotency markers and a genome-wide increase in DNA methylation [46–49]. In the naive state Xist is stably repressed by the core pluripotency factors Nanog, Oct4 and Sox2, as well as by stem cell factors such as Rex1 and Prdm14 [50]. Indeed, the depletion of Nanog, Oct4 or Rex1 in undifferentiated ESCs results in ectopic Xist upregulation [36, 51–54]. Downregulation of these factors at the onset of differentiation allows Xist upregulation. Thus, pluripotency factors are important XCI regulators that control developmental timing of XCI. Interestingly, however, this is not a one-way regulation: Double X-dosage also increases pluripotency factor expression and reduces global DNA methylation in females (Fig 1.2) [55–57]. Female mESCs with two Xas therefore differentiate more slowly than XO or XY cells, and this delay is overcome when XCI is induced [58]. An interesting perspective is that these global X-dosage dependent differences might contribute to proper XCI regulation via two distinct effects: On the one hand, globally higher methylation levels in males are also reflected in a more strongly methylated Xist promoter in males than in females [59, 60], and this promoter methylation has been shown to be essential for Xist repression [61, 62]. Possibly, lower Xist promoter methylation levels thus create a female-specific window of opportunity for Xist upregulation [58, 63]. On the other hand, double X-dosage increases the levels of Xist repressing pluripotency factors. One might speculate that this higher expression of pluripotency factors is required to ensure proper timing of Xist upregulation: Their repressive effect could slow down Xist upregulation, thereby ensuring that Xist is generally upregulated from one X chromosome at a time, thus preventing biallelic Xist upregulation [63]. In summary, cells thus encounter a checkpoint that ensures that development is arrested until proper dosage compensation is achieved. Global sex-specific differences in methylation levels and pluripotency factor expression could contribute to accurate female-specific and monoallelic XCI (Fig 1.2).

1.4 Mechanisms of gene silencing

How exactly a single lncRNA can orchestrate the silencing of a complete chromosome is an enigma that has only recently started to be unraveled as pull-down studies and genetic screens

for factors required for silencing have identified more and more Xist interaction partners [15, 64–67]. It is becoming increasingly clear that the Xist RNA recruits members of different gene silencing pathways, with the most essential ones being Spen, which in turn recruits transcriptional corepressors and histone deacetylases, and the polycomb complexes, which are recruited by hnRNPK [68, 69]. Possibly, there are groups of genes that are more susceptible to one or the other mechanism of silencing [70]. Several other factors such as the m6A methylation machinery and the Lamin B receptor have also been identified as Xist interaction partners and proposed to play a role in silencing [69, 71, 72]. Most of these proteins are recruited by specific repeat domains of the Xist RNA, such as Repeat A which recruits Spen. Deletion of Repeat A strongly abolishes silencing, and, interestingly, also Xist upregulation, possibly indicating some feedback of Xist-mediated silencing on Xist upregulation [12, 73]. In line with this, the deletion of Spen in female mESCs also prevents Xist upregulation supporting the notion that efficient Xist RNA-mediated silencing is crucial for correct XCI onset [74].

To guarantee stable maintenance of the silent Xi, multiple mechanisms likely act together, one of the most important ones being DNA methylation. Promoters on the Xi that contain CpG islands are highly methylated [75–77] and loss of DNA methyltransferases (Dnmts) by inhibition or KO causes partial X-chromosome reactivation [78, 79]. In the extra-embryonic lineages polycomb also seems to be essential for the maintenance of the silent state, as the loss of PRC2 components results in partial reactivation of the Xi [80]. These mechanisms likely complement one another to ensure reliable propagation of the silent state. Recent advances also suggest that phase separation could help in the establishment of the repressive compartment, with lower concentrations of PolII and TFs within the phase-separated repressive compartment [81]. The Xist E Repeat has been proposed to recruit different RNA binding proteins (PTBP1, MATR3, TDP-43 and CELF1) that seed the condensate on the Xi which then maintains silencing and anchors Xist to the Xi territory [81]. This could also contribute to explaining how Xist RNA binds to chromatin and how its binding is restricted in *cis* - a question that has occupied the field for many years. Anchoring of Xist RNA to chromatin is not specific to the X chromosomes, as autosomes expressing Xist transgenes can also be coated by Xist RNA [82]. Studies that have deleted or blocked different regions of the Xist RNA suggest that Xist RNA localization is mediated by a combination of different domains, including repeats E and C [12, 83–87].

Major progress has also been made in understanding what determines the progression of Xist coating and silencing along the X chromosome. High-resolution mapping of Xist RNA on chromatin has revealed that those parts of the X chromosome that are closest to the *Xist* locus in 3D space are the earliest sites covered by Xist [88].

1.5 Escape from XCI

While some X-linked genes are silenced rapidly, others take longer to be shut down. The most extreme case, however, are loci that are able to escape gene silencing completely. In humans, around 15% of genes escape constitutively, and another 10% escape silencing to varying degrees in different tissues and individuals [89]. In mice, the fraction of escape genes is considerably lower but still they exist [90, 91]. Allele-specific single cell RNA-Seq analysis even suggests that in mice genetic variation between different strains can affect the efficiency and the dynamics

of gene silencing. In the most extreme case certain genes might be silenced in one genetic background and escape silencing in another. Also within the *Xist* gene, polymorphisms might affect silencing efficiency, for instance if they lie in regions of the RNA that interact with silencing factors [92]. Escape genes might also contribute to the phenotype associated with X-chromosome aneuploidies, for instance if they have a Y homolog or if they control female-specific phenotypes (see section 1.7) [93, 94]. Still, little is understood about how these escape genes maintain a transcriptionally active state in the repressed Xi environment. Possibly 3D structure could play a role because escape genes tend to have higher interaction frequencies with each other, though this increase in interaction might also be consequence rather than cause of their transcriptional activity [6, 14, 70].

1.6 Species-specific differences in XCI

The frequency of escape is not the only difference in the XCI process between species. The most distinguishing characteristic of mice surely is that they have also evolved an imprinted form of XCI preceding the random wave, in which all cells inactivate the paternal X chromosome (Fig 1.3a). Recently, this imprint has been shown to be mediated by trimethylation of Lysine 27 on histone H3 (H3K27me3), which covers a ∼450 kb region upstream of *Xist* and prevents Xist upregulation from the maternal allele [95]. In the late blastocyst, the cells of the inner cell mass erase this imprint and reactivate the paternal X chromosome. Subsequent downregulation of pluripotency factors derepresses Xist, initiating a second round of XCI, this time random (Fig 1.3a). The observation that mice are the only species with an imprinted form of XCI could be explained by a difference in developmental timing: Maternal to zygotic transition in gene expression occurs much earlier in mice than in other species and could result in an earlier need for dosage compensation [96–98].

Even when comparing the random wave of XCI between species, profound differences in the patterns of Xist upregulation become apparent (Fig 1.3b). While mice were until recently thought to strictly upregulate Xist in a monoallelic fashion, rabbit embryos initially express Xist transiently from both X chromosomes in the majority of cells [99]. Even more extreme are human embryos: Initially all cells upregulate Xist biallelically and, in contrast to rabbits, stably maintain the biallelic expression over a period of several days, although without inducing complete gene silencing [99, 100]. In vivo, this biallelic Xist expression is only resolved after implantation into the uterus, which complicates its investigation. Until recently, an in vitro model system that allowed to observe the onset of random XCI in human was lacking. Novel advances in human ESC culture have now allowed researchers for the first time to derive naive hESCs that undergo XCI with a random choice of the Xi upon differentiation and can therefore serve as a valuable human model for XCI [101]. The availability of such a model system will hopefully rapidly advance our understanding of how the XaXi state is established in humans.

Despite these differences during the onset, XCI seems to result in the same outcome across species, as somatic cells of all mammals studied so far, express Xist from one out of two X chromosomes. In diploid female cells, XCI thus consistently results in one inactive and one active X chromosome.

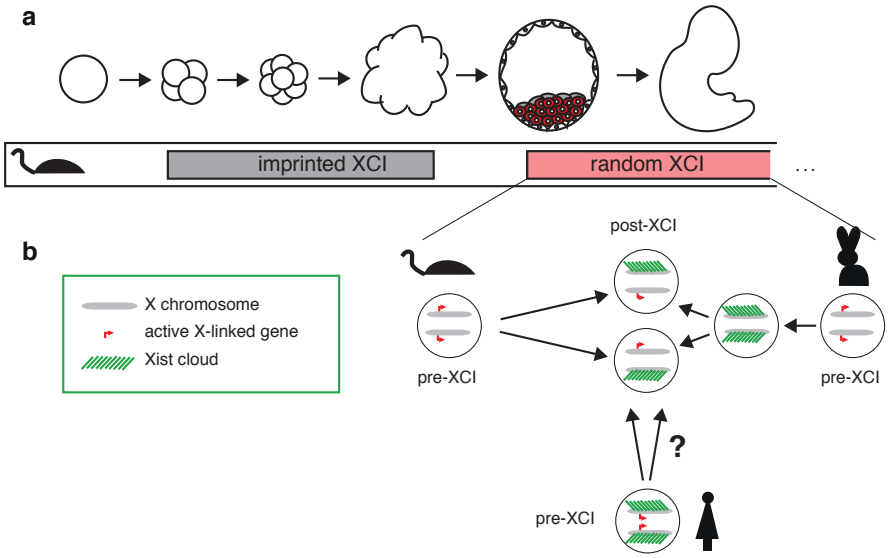

FIGURE 1.3: Regulation of X-chromosome inactivation during early mouse, rabbit and human development. a) Mice are the only species that undergo an imprinted wave of XCI, which initiates in the 2-4 cell stage and results in inactivation of the paternal X chromosome in all cells. Imprinted XCI is reversed in the inner cell mass of the blastocyst around E4.5. Subsequently, random XCI occurs. b) In mouse, rabbit and human, random XCI results in the inactivation of one randomly chosen X chromosome (Xp or Xm). However, the route towards the XaXi state is different. Left: In mice, most cells directly upregulate Xist monoallelically. Right: In rabbits, most cells go through a transient period of biallelic Xist expression. Bottom: Human embryos stably maintain biallelic XIST expression over several days without inducing complete gene silencing. How they transit to the XaXi state is still unclear.

1.7 XCI patterns in X aneuploidies and polyploid cells

X-chromosome aneuploidies cause severe phenotypes. Turner syndrome (45,X) for instance results in physical abnormalities and ovarian dysgenesis, and most 45,X embryos die *in utero* [93, 102]. The inactivation patterns of cells with X aneuploidies can help to shed light on the rules of XCI: Females with Turner syndrome (XO) do not undergo XCI, while male Klinefelter patients (XXY) will inactivate one of their two X chromosomes [103]. Thus, the decision of whether XCI is initiated depends only on the number of X chromosomes, and is not influenced by the presence or absence of a Y chromosome. Females with more than two X chromosomes inactivate all but one X [104], while tetraploid embryos and ESCs will maintain two active X and triploid embryos are a mixture of cells with one and two Xas [105–108]. These observations indicate that one active X chromosome is maintained per diploid set of autosomes, and suggest that cells have a way to assess their number of X chromosomes in relation to their ploidy.

1.8 The regulatory principles governing XCI onset

In her 1971 paper Lyon goes on to ask: "*Thus in considering the mechanism of X inactivation there are two distinct problems: (1) how is the differentiation* [between the two X chromosomes] *brought about, and (2) how is it maintained through subsequent cell divisions and concomitant chromosome replications?*" [109].

Indeed, it is one of the most fascinating aspects of XCI that two functionally and genetically

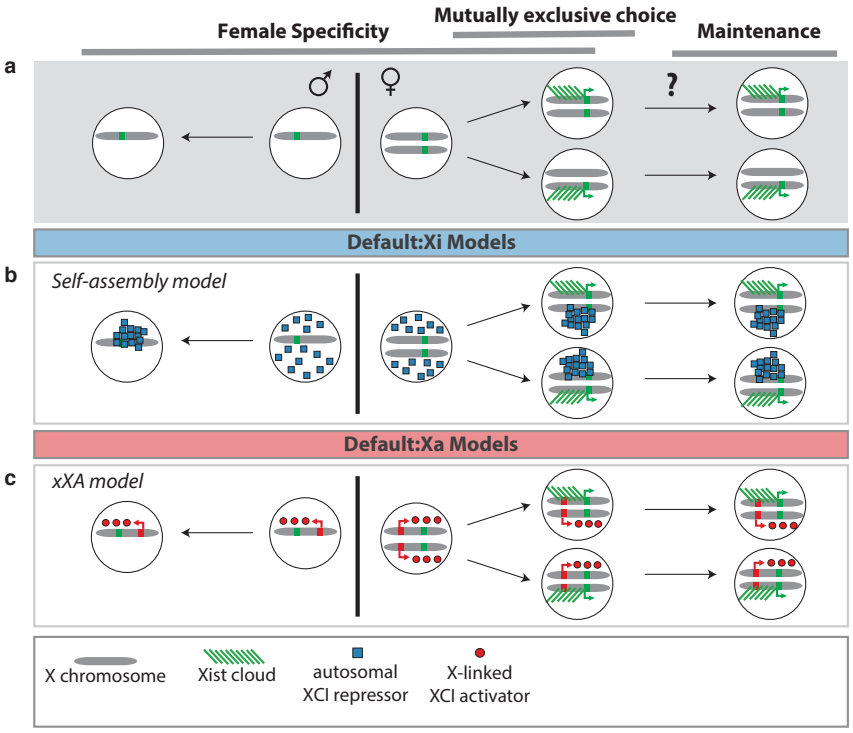

FIGURE 1.4: Alternative XCI models. a) Female specificity: XCI is only initiated in female cells with two X chromosomes (right), but never in male cells with a single X (left); mutually exclusive choice: Xist is monoallelically upregulated from one of the two X chromosomes; maintenance: The decision for the Xa and Xi is then stably maintained throughout cell division. b) Default:Xi models assume inactivation as the default fate for each X chromosome. Female specificity is ensured through a protective factor that self-assembles into a single cluster and protects exactly one X chromosome from inactivation. c) Default:Xa models assume activity as the default fate for each X chromosome, and XCI is initiated only if at least two X chromosomes are present. X dosage could be sensed by a diffusible *trans*-acting X-encoded factor (tXA) that is present in a double dose in female cells and activates XCI dose-dependently. Upon XCI Xist will silence the *trans*-acting tXA to prevent Xist upregulation from the other allele.

equivalent X chromosomes within the same nucleus undergo completely different fates that are then stably maintained for all further cell divisions. To achieve this, a cell must first assess how many X chromosomes it possesses and only initiate XCI if it has more than one X per diploid set of autosomes (female specificity). Each cell must then choose one (or more) X to inactivate (Xi) and one X to stay active (Xa) (mutually exclusive choice). Once this decision has been made, it must be stably maintained (maintenance) (Fig 1.4a) [63]. In the 60 years since Lyon had first proposed the process of XCI, numerous regulatory concepts have been developed to answer these questions. Below, I will attempt an overview of existing hypotheses on the regulation of XCI onset.

1.8.1 Female specificity

How do cells count their number of X chromosomes to ensure that XCI is robustly initiated if they have two (or more) X chromosomes, but never if they only possess a single X chromosome? In essence, two alternative approaches are possible: Each X chromosome could get inactivated by default (Default:Xi), which would mean that in each cell, male or female, a single X would need to be protected from inactivation. In the alternative scenario, each X chromosome stays active by default (Default:Xa) and cells will need to sense the presence of more than one X

chromosome to trigger inactivation [63].

Default:Xi
Many early XCI models have proposed a protective factor that exists at exactly one copy per cell. This factor could be in the form of a nuclear attachment site protecting the attached X from inactivation [110], or in the form of a single molecule RNA or protein "blocking factor" that will bind either one or the other X [7, 20, 109, 111]. Advanced models have addressed an essential draw-back of this hypothesis, the impossibility of reliably maintaining exactly one RNA or protein molecule per cell, by suggesting the protective factor to self-assemble into a single large cluster, that will, again, associate with either one or the other X (Fig 1.4b) [112, 113]. However, it remains challenging to explain how tetraploid cells would form exactly two stable protective clusters, or how triploid cells would sometimes form one, and other times two (but never three) stable clusters.

Default:Xa
In the default:Xa scenario, cells must sense the presence of more than one X chromosome to initiate XCI. This sensing has been proposed to occur by pairing of homologous regions on the two X chromosomes [114–116]. However, recent data show that preventing pairing by tethering one or both X to the nuclear lamina has no effect on initiation of XCI [117]. Alternatively, sensing could be achieved by a diffusible factor (RNA or protein) that is produced from the X chromosome and thus exists at roughly two-fold higher levels in females than in males (red in Fig 1.4d) [109]. Let us call this factor a *trans*-acting XCI activator (tXA). Indeed, heterokaryons with an XX and XY nucleus in the same cytoplasm inactivate any of the three X chromosomes with equal probability, suggesting that they all receive the same XCI signals [43]. Additional experimental support for a dose-dependent X-linked XCI activator comes from the observation that additional X chromosomes increase the rate with which XCI is initiated [107]. Importantly, in the tXA model, cells must convert a quantitative signal (1x or 2x tXA copies) into a qualitative response (no XCI or XCI): XCI must be initiated in a switch-like manner only if tXA exceeds a certain threshold concentration that lies somewhere between the dose produced from one and two X chromosomes, thus somewhere between male and female cells. Such ultrasensitivity can for instance arise from molecular titration, cooperative binding or positive feedback [118–121].

1.8.2 Mutually exclusive choice for one Xi

How do the two X chromosomes that are functionally and genetically equivalent before the onset of XCI, assume opposing transcription states? Default:Xi models propose the association of a single molecule or a cluster of molecules with either one or the other X chromosome, thereby preventing XCI on the protected X chromosome (self-assembly model, Fig 1.4b). In the Default:Xa model, tXA silencing upon XCI could prevent inactivation of the second X chromosome (Fig 1.4c), as already proposed by Mary Lyon and others [107, 109].

Interestingly, other biological regulations often solve problems of mutual exclusivity by mutual inhibition between two entities in a so called toggle switch [122]. The toggle can exist in two alternative stable states: Either entity 1 is active and represses entity 2 or vice versa. For instance, mutual inhibition between two lineage-determining transcription factors often governs

cell fate decisions [123, 124]. But also other decision-making processes such as the phage lambda lysis/lysogeny decision can be explained by a toggle switch of mutual repression between the two transcription factors CI and Cro which repress each other's production in an ultrasensitive manner and promote or repress the lysogenic state, respectively [125–128]. The above-discussed self-assembly model would in principle generate exactly such a scenario of pure mutual inhibition between the two X chromosomes, because each X would sequester the protective factor from the other X. However, this model still reaches its limits when it comes to polyploid cells. It is, thus, not immediately apparent how such mutual inhibition could be implemented between the two X chromosomes.

1.8.3 Stable XaXi maintenance

In the early stages, silencing of the Xi requires the continuous presence of Xist RNA in *cis*. Therefore, the alternative allelic Xist expression states on Xa and Xi must be memorized. One of the earliest XCI models suggested the insertion of a DNA element into the active X, as an irreversible event that would stably maintain the different expression states [129]. In light of large-scale sequencing data that fail to detect such Xa-specific DNA insertions and in light of the fact that XCI is reversed in the germ line, this hypothesis now seems rather implausible. Without changes to the DNA sequence, how can the alternative Xist expression states then be stably maintained on the two genetically identical Xist alleles within the same nucleus? This task reminds one of the problem that all multicellular life faces: How can different cell fates be encoded by identical DNA sequences? An important difference to the memory of cell fate decisions is that during XCI not only a cellular but even an allelic memory must be generated, to allow the stable maintenance of two alternative expression states within the same nucleus. In other biological contexts such *cis* memory is often maintained by epigenetic mechanisms, such as DNA methylation or positive feedback in nucleosome modifications [130–133]. *Cis* memory could also be generated by a physical mechanism, such as phase separation where the Xist RNA triggers the formation of a protein-condensate that eventually becomes self-sustained [81].

1.9 Aim of the study

Although several concepts to explain parts of XCI initiation have been developed, very few have been formalized theoretically and none can explain all aspects of XCI onset. Similarly, several regulators governing the initiation of random XCI are known today, but their precise interplay still remains elusive. In this work, we approach XCI onset from an alternative angle, by using mathematical modeling to identify the minimal XCI network and suggest mechanisms that could mediate the predicted regulatory loops. The goal of this work is not to capture all possible details of the XCI process nor to precisely determine the values of all model parameters, rather it is to identify the unknown principles that could govern XCI onset.

2 Results

2.1 Systematic identification of a minimal model structure for XCI

To identify the core regulatory interactions necessary for female-specific and monoallelic Xist expression, we systematically compared alternative regulatory networks of equal complexity. The XCI regulatory network must stabilize the XaXi state and destabilize the XiXi state in female cells to ensure stable monoallelic Xist expression. In addition, it must destabilize the Xi state with high Xist expression ($Xist^{high}$) and stabilize the Xa state with low Xist expression ($Xist^{low}$) in male cells to ensure female specificity. XCI likely involves at least one X-linked regulator, that confers information about the number of X chromosomes that a cell possesses. To understand which regulators were minimally required, we classified all X-linked Xist regulators into eight categories based on three different features: Any Xist regulator can either activate (A) or repress (R) Xist. It can either act in *cis* (c), only affecting the Xist allele on the same chromosome, or it can act in *trans* (t), regulating both Xist alleles. Lastly, it can either be silenced upon Xist upregulation from the same chromosome, or it might escape silencing (e), as observed for some X-linked genes (Fig 2.1). In fact, other *cis*-regulatory elements such as enhancers are conceptually similar to the category of escaping XCI regulators as their function does not necessarily depend on the act of transcription.

To systematically test which regulator types were minimally required we constructed ordinary differential equation (ODE) models of a female cell with two X chromosomes that contain Xist and either a single regulator or two regulator types at a time (eight single-regulator models (Table 2.1) and 28 two-regulator models (Table 2.2)).[1]

[1]The ODE models were constructed and analyzed by Edda Schulz.

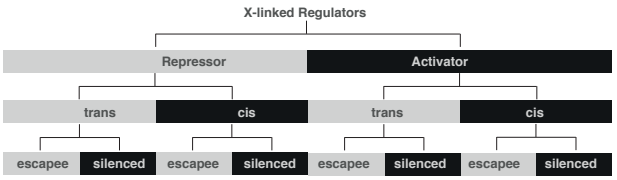

FIGURE 2.1: Classification of all X-linked regulators into eight categories, depending on whether they activate or repress Xist, whether they act in *cis* or *trans*, and whether they are silenced or escape silencing.

Table 2.1: Single-regulator ODE models

	Chr. X1	**Chr. X2**	
Xist	$\frac{dx_1}{dt} = f(r) - x_1$	$\frac{dx_2}{dt} = f(r) - x_2$	(1)
cis regulator r	$f(r) = a + b\frac{r_1^n}{r_1^n + K^n}$	$f(r) = a + b\frac{r_2^n}{r_2^n + K^n}$	
trans regulator r	$f(r) = a + b\frac{(0.5\cdot(r_1+r_2))^n}{(0.5\cdot(r_1+r_2))^n + K^n}$		
	XA: a=0, b=1 XR: a=1,b=-1		
Regulator			(2)
silenced	$\frac{dr_1}{dt} = 1 - \frac{x_1^n}{x_1^n + K^n} - r_1$	$\frac{dr_2}{dt} = 1 - \frac{x_2^n}{x_2^n + K^n} - r2$	
escaping	$\frac{dr_1}{dt} = 1 - r_1$	$\frac{dr_2}{dt} = 1 - r_2$	

Table 2.2: Two-regulator ODE models

	Chr. X1	**Chr. X2**	
Xist	$\frac{dx_1}{dt} = f(r_A)\cdot f(r_B) - x_1$	$\frac{dx_2}{dt} = f(r_A)\cdot f(r_B) - x_2$	(3)

The models describe the transcription and degradation dynamics of Xist and the respective regulator(s) and how they interact through mutual regulation of their transcription rates on a population averaged level. An Xist regulator either positively (A) or negatively (R) affects Xist production. Xist negatively affects the production of silenced regulators on the same chromosome. Each regulator is produced with a maximal production rate of 1 and degraded with a degradation rate of 1 so that the levels of all regulators are scaled between 0 and 1. All regulatory functions are modelled with Hill equations, a standard model for biological interactions with sigmoidal responses, as exhibited by ultrasensitive or cooperative regulatory processes. It has two parameters: The Hill coefficient n determines the level of cooperativity, and the threshold K determines at what level of input the output is half-maximal. We were thus able to test different degrees of non-linearity and different efficiencies of repression or activation for each of the regulatory interactions ($1 \leq n \leq 5$ and $0.01 \leq K \leq 10$). Each network was simulated with a large number of randomly sampled parameter sets (>10,000 per model) with different regulatory strength to test whether a given network architecture could in principle produce the desired characteristics.

We started by simulating the XaXi state in female cells (Simulation 1, Fig 2.2a). With this analysis we sought to investigate the post-XCI state where female cells have already established Xa and Xi. We therefore initiated one of the two X chromosomes in the Xa state, where Xist expression is low and regulator expression is maximal, and the other X in the Xi state, where Xist expression is high and expression of all silenced regulators is low, while escaping regulators

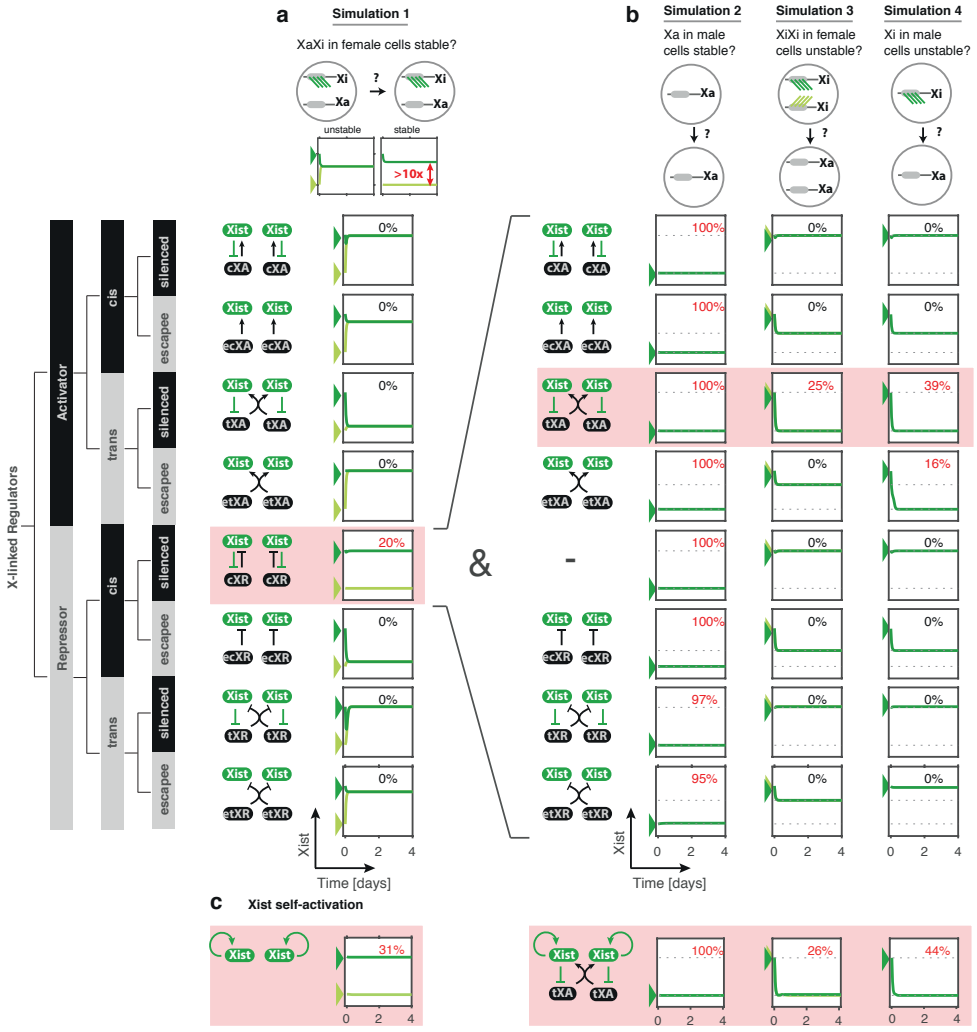

FIGURE 2.2: Comparison of alternative model structures. Each network was translated into an ODE model, describing two X chromosomes, each carrying Xist and the respective regulator(s). a) Simulation of XaXi maintenance in the eight ODE models containing Xist and a single regulator (see schemes). Each network was simulated with >10,000 randomly chosen parameter sets. The simulation was initiated from the XaXi state, where Xist is transcribed from the Xi (dark green), but not from the Xa (light green). Per network one example simulation is shown and the percentage of parameter sets that could maintain the XaXi state is indicated. b) The cXR, the only regulator that could maintain the XaXi state in a), was separately combined with one other regulator at a time, forming seven additional two-regulator models. Three additional simulations were performed with all parameter sets that could maintain the XaXi state. Simulation 2 tests whether the Xa state could be maintained in male cells, simulation 3 tests whether the XiXi state could be destabilized in female cells, and simulation 4 tests whether the Xi state could be destabilized in male cells. c) A pure positive feedback can also stabilize alternative Xist expression states. Left: Simulation 1 with an ODE model, in which mutual repression between Xist and cXR has been replaced by a self-activating feedback of Xist. Right: Simulation 2-4 with a network combining the positive feedback of Xist with a tXA. Arrowheads indicate the initial conditions, and dotted lines denote the Xa and Xi state of the respective parameter set in simulation 1. Red shading indicates the models that could fulfill the requirements.

are unaffected.

Of the single-regulator models only the cXR model could stably maintain the alternative Xist expression states on Xa and Xi in a fraction of parameter sets (Fig 2.2a). Also among the two-regulator models, all models containing cXR, but none of the other models, could stabilize the XaXi state (Table 2.3). Thus, cXR is necessary and sufficient for stable XaXi maintenance.

Table 2.3: Percentage of monoallelic parameter sets

	cXA	**ecXA**	**tXA**	**etXA**	**cXR**	**ecXR**	**tXR**	**etXR**
cXA	0	0	0	0	10.2	0	0	0
ecXA		0	0	0	14.6	0	0	0
tXA			0	0	14.4	0	0	0
etXA				0	14.9	0	0	0
cXR					20.0	8.1	9.4	8.8
ecXR						0	0	0
tXR							0	0
etXR								0

In a next step, we tested all models that passed the XaXi criterium for their ability to stabilize the Xa state in male cells (Simulation 2, Fig 2.2b), and to destabilize the XiXi state in females (Simulation 3, Fig 2.2b) and the Xi state in males (Simulation 4, Fig 2.2b). To simulate male cells, the models were reduced to containing a single X chromosome with one copy of Xist and each regulator. While all candidate models were able to stably maintain the $Xist^{low}$ state in males (Simulation 2, Fig 2.2b), a *trans*-acting Xist activator (tXA) was necessary to destablize the Xi state in males and the XiXi state in females (Simulation 3 & 4, Fig 2.2b). The tXA factor did not necessarily have to be silenced to ensure destabilization of the $Xist^{high}$ state in males (Simulation 4), but it had to be silenced to prevent stable biallelic Xist expression in females (Simulation 2), consistent with Mary Lyon's early considerations.

With a systematic network comparison, we have thus identified the single simplest network that stabilizes the XaXi (and the Xa in males) and destabilizes the XiXi (and the Xi in males). It consists of two silenced regulators, a cXR and a tXA. The regulatory circuits implemented by these two regulators are a *trans*-acting negative feedback through tXA silencing, and a double negative feedback of mutual repression between Xist and cXR.

To demonstrate that the double negative feedback could in theory also be substituted by a purely positive self-reinforcing feedback that is independent of the cXR, we also tested a model in which Xist activates its own expression in a purely positive feedback, and found that this can also recapitulate XaXi stabilization, and together with tXA, XiXi destabilization (Fig 2.2c):

$$\frac{dx_1}{dt} = \frac{x_1^{n_a}}{x_1^{n_a} + K_a^{n_a}} \cdot \frac{(0.5 \cdot (tXA_1 + tXA_2))^{n_b}}{(0.5 \cdot (tXA_1 + tXA_2))^{n_b} + K_b^{n_b}} - x_1$$

In conclusion, our systematic network comparison identified the previously predicted tXA factor as necessary to ensure female specificity and prevent biallelic Xist expression. It also makes the novel prediction that a self-reinforcing local feedback is required to stabilize the XaXi state, and can be implemented by mutual repression between Xist and cXR.

TABLE 2.4: Equations of the stochastic cXR-tXA model

	Chr. X1	**Chr. X2**	
Xist	$\frac{dx_1}{dt} = p21 \cdot f(cXR_1) \cdot f(tXA) - 0.1733 \cdot x_1$	$\frac{dx_2}{dt} = p21 \cdot f(cXR_2) \cdot f(tXA) - 0.1733 \cdot x_2$	(6)
	$f(cXR_1) = 1 - \frac{cXR_1^{p13}}{cXR_1^{p13} + (p_{22} \cdot p_{14})^{p13}}$	$f(cXR_2) = 1 - \frac{cXR_2^{p13}}{cXR_2^{p13} + (p_{22} \cdot p_{14})^{p13}}$	
	$f(tXA) = \frac{(0.5 \cdot (tXA_1 + tXA_2))^{p11}}{(0.5 \cdot (tXA_1 + tXA_2))^{p11} + (p_{23} \cdot p_{12})^{p11}}$		
tXA	$\frac{dtXA_1}{dt} = p_{23} \cdot (1 - \frac{x_1^{p3}}{x_1^{p3} + (p_{21} \cdot p_4)^{p3}}) - tXA_1$	$\frac{dtXA_2}{dt} = p_{23} \cdot (1 - \frac{x_2^{p3}}{x_2^{p3} + (p_{21} \cdot p_4)^{p3}}) - tXA_2$	(7)
cXR	$\frac{dcXR_1}{dt} = p_{22} \cdot (1 - \frac{x_1^{p5}}{x_1^{p5} + (p_{21} \cdot p_6)^{p5}}) - cXR_1$	$\frac{dcXR_2}{dt} = p_{22} \cdot (1 - \frac{x_2^{p5}}{x_2^{p5} + (p_{21} \cdot p_6)^{p5}}) - cXR_2$	(8)

2.2 XaXa → XaXi: cXR-tXA model can reproduce XCI onset

We next asked whether the cXR-tXA model (Fig 2.3a) can also explain the initial establishment of the XaXi state at the onset of XCI, where female cells transit from a state with two active X (XaXa) to a state with one active and one inactive X (XaXi), making a mutually exclusive choice for an inactive X chromosome (Fig 2.3b).

The XaXa → XaXi transition, where Xist is randomly upregulated from one of the two X chromosomes, requires a symmetry break between the two X chromosomes, mediated by stochastic fluctuations in Xist or its regulators. To account for this, we developed a stochastic implementation of the cXR-tXA network. This necessitated the implementation of a few changes compared to the ODE models of the previous section:

- Absolute molecule counts
  In the stochastic simulation absolute molecule counts are important as they will affect the variability between alleles. Some regulators might be present at low levels, such that random fluctuations in their expression level will contribute strongly to allelic variability. To account for this, we added scaling factors to the production terms such that the regulator levels would vary within biologically relevant ranges.

- Silencing kinetics
  In the previous section we analyzed the state of post-XCI cells, where the kinetics of gene silencing are not important because they do not affect the steady state in which the regulators are already silenced on the Xi. Here, however, they are important because they can impact the dynamics with which Xist is upregulated. As the silencing speed is variable for different genes and the identity of the regulators is unknown we added two additional parameters sil_{cXR} (=p_7) and sil_{tXA} (=p_8), that describe how fast after Xist upregulation tXA and cXR are silenced, respectively. We assumed that the Xist RNA must transit through sil_{tXA} or sil_{cXR} intermediate states before it becomes competent to silence tXA

or cXR. These transitions occur with rate 1 h^{-1} such that the number of intermediate states given by the parameters is equal to the mean silencing delay in hours ($\in$ [1 h, 20 h]).

- Xist RNA stability
  We adapted the Xist RNA degradation rate to experimental measurements which determine the Xist RNA half-life to be between 2 h and 6 h [85, 134]. We therefore used the mean of 4 h, which results in a degradation rate of 0.1733 h^{-1} (=$\ln(2)/t_{1/2}$).

- tXA and cXR stability
  Degradation rates for cXR and tXA were kept at 1 h^{-1} as their molecular identity remains unknown and therefore no experimental estimates of their stability exist but the unknown degradation kinetics are indirectly accounted for by assuming that they modulate the kinetics of tXA and cXR silencing.

The adapted formulation of the model for the stochastic simulations is summarized in Table 2.4. Table 2.5 summarizes the tested parameter ranges of the stochastic cXR-tXA model.

Table 2.5: Parameters of the stochastic cXR-tXA model

Parameter	**Function**	**Value**
p_3	Xist –| tXA, Hill coefficient n	1 ... 5
p_4	Xist –| tXA, threshold K	0.01 ... 10 (log distributed)
p_5	Xist –| cXR, Hill coefficient n	1 ... 5
p_6	Xist –| cXR, threshold K	0.01 ... 10 (log distributed)
p_7	silencing delay cXR	1 ... 20 h
p_8	silencing delay tXA	1 ... 20 h
p_{11}	tXA $\rightarrow$ Xist, Hill coefficient n	1 ... 5
p_{12}	tXA $\rightarrow$ Xist, threshold K	0.01 ... 10 (log distributed)
p_{13}	cXR –| Xist, Hill coefficient n	1 ... 5
p_{14}	cXR –| Xist, threshold K	0.01 ... 10 (log distributed)
p_{18}	transition rate between silencing intermediates	1 h^{-1}
p_{21}	scaling factor Xist	50 ... 500 (log distributed)
p_{22}	scaling factor cXR	50 ... 500 (log distributed)
p_{23}	scaling factor tXA	50 ... 500 (log distributed)

Each parameter set that could maintain the XaXi (and Xa) state, and destabilize the XiXi (and Xi) state in the ODE simulations in the previous section was combined with ten sets of randomly sampled silencing delays and scaling factors. To perform the simulations we used the Gillespie algorithm which numerically simulates the time evolution of a chemical reaction system, taking into account inherent fluctuations that are ignored in the ODE formulation, which assumes the time-evolution to be continuous and deterministic [135, 136]. The Gillespie algorithm simulates individual trajectories of single cells. We simulated the system for 100 h with both Xist alleles starting from the off state (XaXa). At each hour of the simulation, each chromosome was classified as Xist+ if Xist levels exceeded 20% of the $Xist^{high}$ state in the ODE analysis, and otherwise as Xist-. For a subset of parameters the model can indeed reproduce robust monoallelic upregulation of Xist, with a high fraction of cells ending in the XaXi state (Fig 2.3c). For one example parameter set the Xist trajectories in two individual cells and the fraction of mono- and biallelic cells are shown in Fig 2.3d. The cXR-tXA model can thus explain the random and mutually exclusive choice for an Xi.

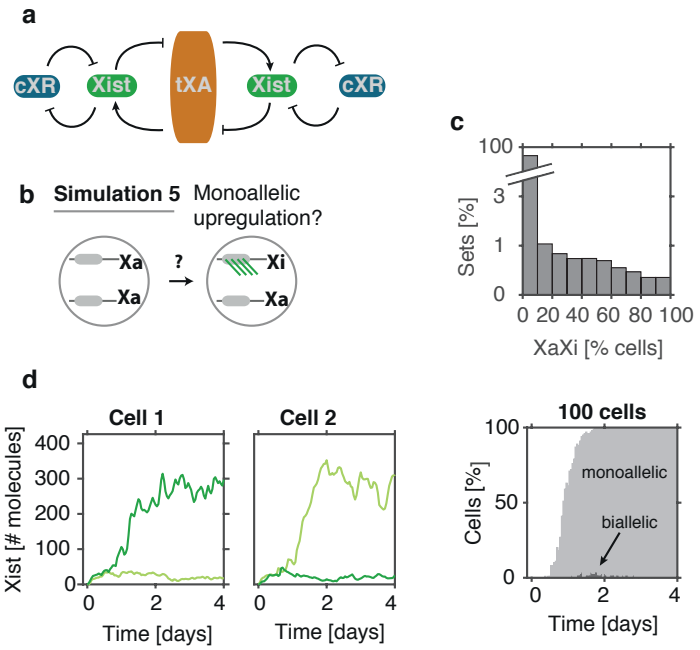

FIGURE 2.3: The cXR-tXA model can recapitulate monoallelic Xist upregulation. a) Scheme of the cXR-tXA model. b) Scheme of the stochastic simulation for monoallelic Xist upregulation. Both X chromosomes start from the Xa state (XaXa), one X chromosome should upregulate Xist (XaXi). c) Distribution of the percentage of cells that end in the XaXi state in all parameter sets. d) Example simulation with a monoallelic parameter set. Left: single cell trajectories (light green: Xist1, dark green: Xist2). Right: Fraction of monoallelic (light grey) and biallelic (dark grey) Xist expressing cells in a population of 100 simulated cells.

To understand how monoallelic Xist upregulation is achieved in the cXR-tXA model we estimated the allelic Xist steady state levels for different tXA concentrations in a monoallelic parameter set (Fig 2.4a, top). To analyze the allelic steady states, we performed an ODE simulation as in section 2.1 starting from different initial conditions of Xist and cXR (Xist = 0, 0.1, 0.2, ...1; cXR = 1- Xist). The tXA concentration was kept constant during the simulation ($tXA \neq f(Xist)$), but the input tXA concentration was varied, reflecting different total steady state tXA doses. This allowed us to investigate, what allelic steady states Xist and cXR will assume for a given tXA dose, and whether these Xist and cXR steady states will differ depending on the initial conditions.

The top panel in Fig 2.4a shows the allelic Xist steady states reached after 100 h of simulation. The tXA level is scaled to the level that would correspond to 0, 1 or 2 active X chromosomes (XiXi, XaXi, XaXa), as calculated from the $Xist^{high}$ and $Xist^{low}$ steady states in the original ODE simulation (section 2.1). In the presence of a single tXA dose the system exhibits bistability where both, the low Xist expression state on the Xa and the high expression state on the Xi, are stable. Thus, post-XCI cells that have already silenced one tXA allele, can stably maintain Xa and Xi. In the presence of a double tXA dose however, only the high expression state is stable, meaning that in a pre-XCI cell with two active X chromosomes (=2x tXA), eventually Xist upregulation will be triggered. Importantly, a cell without any tXA cannot maintain Xist expression, meaning that a female cell which has upregulated Xist biallelically and has silenced both its tXA alleles, cannot maintain Xist expression which destabilizes the XiXi state. This

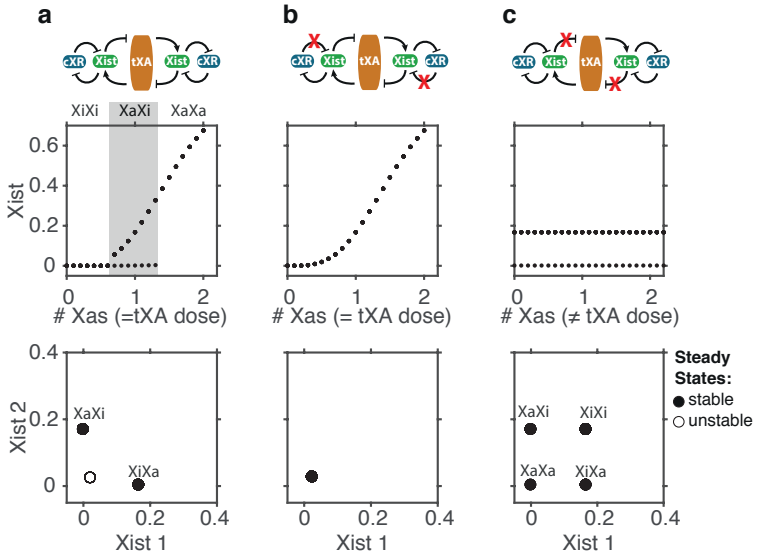

FIGURE 2.4: Role of cXR and tXA. a) Allelic (top) and cellular (bottom) steady states simulated with the ODE model of the full network. To simulate the allelic steady states, the tXA dose was kept constant, and different initial values for Xist and cXR were tested. The Xist expression level after 100 h of simulation is plotted over the number of active X chromosomes (total tXA dose). The shaded area indicates the bistable regime for a single tXA dose corresponding to the monoallelic XaXi state. To simulate the cellular steady states, the simulation was started from different initial values of Xist1 and Xist2 and the cXR1, cXR2, tXA1 and tXA2 initial values were set to their steady state values resulting from the respective Xist1 and Xist2 values. Filled circles indicate stable, open circles unstable steady states. b) Allelic and cellular steady states upon removal of cXR-mediated Xist repression. c) Allelic and cellular steady states upon removal of tXA silencing by Xist. Here, the tXa dose is not affected by Xist, and was set to a constant value (=tXA dose in the XaXi state of the full network).

could constitute a safety net for cells that have accidentally started to inactivate both X chromosomes, as they would switch off Xist again upon complete tXA silencing (more on this in section 2.4 and 2.5). Indeed, we observed cells in the simulation which had initially upregulated Xist biallelically but then resolved this biallelic to monoallelic expression (dark grey in Fig 2.3d). Thus, at the cell level (where the tXA dose can either be 0, 1 or 2) only the monoallelic Xist expression states (XaXi or XiXa) but not the Xist-negative (XaXa) or biallelic (XiXi) expression states are stable (Fig 2.4a, bottom).

To understand the role of the feedbacks mediated by each of the regulators, we separately removed the cXR and tXA modules (Fig 2.4b,c). For cXR we removed the repressive effect it has on Xist by setting the threshold level of cXR required to repress Xist to a very high value (p_{14}=1000). To remove the negative feedback mediated by tXA silencing we set the level of tXA that activates Xist to a constant dose corresponding to the single dose in XaXi cells (since we know that the system is bistable in this regime). We then analyzed the effect of the perturbation on the allelic and cellular steady states: Without cXR the bistability disappears and only a single stable Xist expression state exists per allele which still changes depending on the tXA level (Fig 2.4b, top). Without the tXA regulation, bistability always exists and is independent of the number of active X chromosomes in the cell (Fig 2.4c, top), which makes all cellular states stable because each X chromosome is independent of the fate of the other X chromosome (Fig 2.4c, bottom).

In summary, the positive feedback mediated by mutual repression between Xist and cXR creates

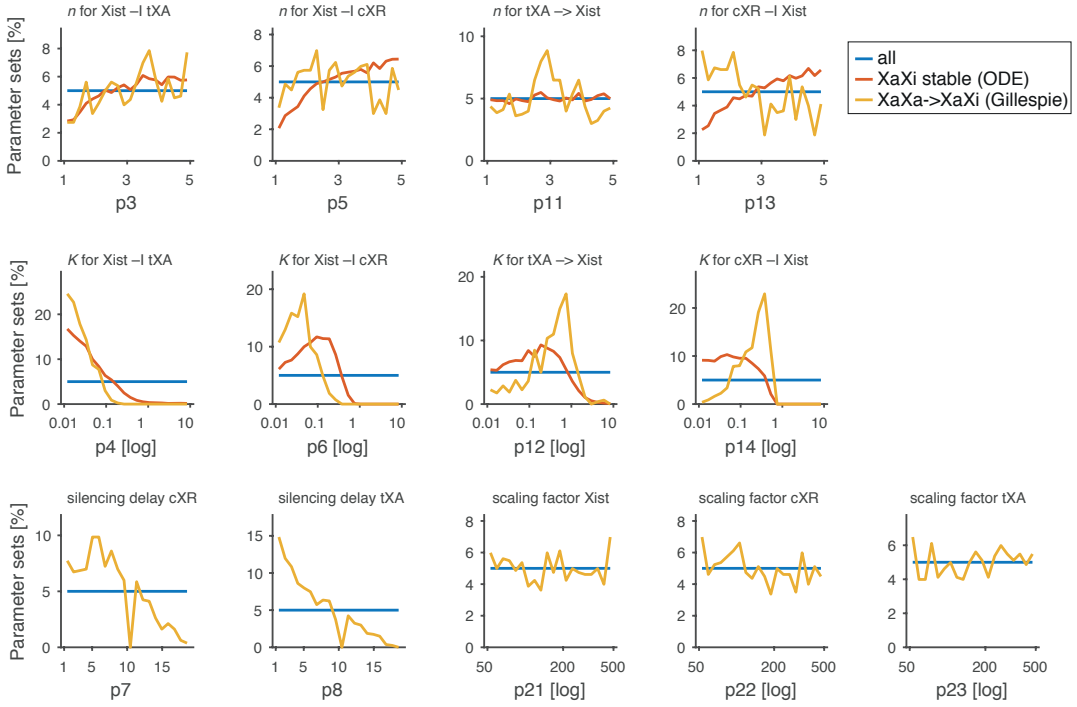

FIGURE 2.5: Distribution of parameter values in all tested parameter sets (blue), the sets that could maintain the XaXi state in the ODE simulation (orange), and the sets that could generate monoallelic Xist upregulation in the stochastic cXR-tXA model (yellow). Silencing delays (p7, p8) and scaling factors (p21, p22, p23) were added in the stochastic model and do not exist in the ODE simulation.

bistability, while the negative feedback mediated by tXA silencing ensures that bistability only exists in cells with 1 Xa and 1 Xi, and is flanked by monostable regions with low Xist expression in the absence of tXA and high Xist expression in the presence of a double tXA dose. The tXA regulation thereby ensures female-specific XCI onset, because only female cells can reach the monostable regime where Xist upregulation is triggered. It also contributes to mutually exclusive choice, because it destabilizes both the XaXa and the XiXi state.

We then asked, which parameters allowed monoallelic Xist upregulation. While there were no trends for the scaling factors (p21, p22, p23), meaning that all values could produce monoallelic Xist upregulation, the silencing delays of cXR and tXA (p7, p8) had to be short so that upon Xist upregulation the decision for the Xi can be quickly locked in by silencing the regulators (Fig 2.5). For the Hill coefficients (p3, p5, p11, p13) no strong trends were observed, except that cooperativity had to be present in the double negative feedback of Xist and cXR repression. The threshold parameters determining how efficiently Xist silences cXR and tXA (p4, p6) are required to be low to ensure efficient silencing. The threshold parameters describing how tXA and cXR regulate Xist (p12, p14) were only restricted by an upper bound for XaXi maintenance. For monoallelic Xist upregulation they are, however, also restricted by a lower bound, likely to allow Xist to be responsive to both in- and decrease in the regulator concentrations.

2.3 cXR-tXA model recapitulates XCI patterns in aneuploid and polyploid cells

As a further test of the cXR-tXA model we wanted to assess whether it can correctly reproduce the XCI patterns in cells with X aneuploidies, which inactivate all but one X chromosome. We selected the 100 parameter sets that could best explain mouse in vivo data on Xist expression (see section 2.4) and used them to simulate diploid cells with 1, 2, 3 or 4 X chromosomes in order to analyze what patterns they predicted in these aneuploid cells. We assumed that each additional X chromosome would produce an additional tXA dose (scheme in Fig 2.6b). Each allele was again classified as Xist+ or Xist- for the last 20 h of the simulation, so that each cell could be categorized as possessing 0, 1, 2, 3 or 4 Xis. Nearly all parameter sets correctly predicted no inactivation in male cells, and 2 Xis or 3 Xis in X tri-, or tetraploidies respectively (Fig 2.6b). This is, again, because the only stable state for the cell remains the one with one Xa. As the tXA stimulus increases further, the driving force for the transition to the $Xist^{high}$ state becomes higher. On the allelic level, the low Xist expression state is instable as long as there is more than one tXA dose (>1 Xa), and therefore Xist upregulation will be triggered until only a single Xa remains (Fig 2.6b).

We then simulated polyploid cells which usually maintain 1 active X per diploid set of autosomes. Thus, while a diploid cell with 4 X chromosomes will inactivate 3 Xs, a tetraploid cell with 4 X chromosomes will only inactivate 2; and triploid cells are effectively a mixture of cells with 1 and 2 active Xs. This suggests that autosomal ploidy in some way regulates XCI outcome. We simulated two alternative scenarios on how ploidy could modulate XCI regulation: In the first one, we assumed that an additional copy of the genome would lead to a doubling of the nuclear volume [137]. In this way, the tXA concentration in tetraploid cells would remain the same as in diploid cells: The doubling of the number of tXA loci would lead to a doubling of the number of tXA molecules, but would be compensated by a two-fold dilution due to the increased nuclear volume (compare Fig 2.6a,b). In the alternative scenario, we assumed that an autosomal factor (blue in Fig 2.6b,c) would repress tXA, for instance through sequestration, such that the effective concentration of active free tXA would be similar in 2n and 4n cells (compare Fig 2.6b,c). As a result of both assumptions, a tetraploid cell with 2 Xas will now reside in the bistable maintenance regime (=2n XaXi, Fig 2.6a,c), while with 4 Xas it will reside in the monostable regime where Xist upregulation is triggered (=2n XaXa, Fig 2.6a,c). Indeed, we found that the majority of tested parameter sets would correctly predict a mixture of cells with 1 and 2 Xis in triploid cells and a majority of cells with 2 Xis in tetraploid cells (Fig 2.6a, c).

However, we found that some of the tested parameter sets tended to inactivate too few or too many X chromosomes (Fig 2.7a,c). Indeed, reaching the biallelic expression state in a tetraploid cell is more difficult than reaching the monoallelic expression state in a diploid cell. A diploid cell only has to discriminate between a 1x and 2x tXA dose (=100% increase). A tetraploid cell, on the other hand, has to be able to distinguish a 2x from a 3x tXA dose (=50% increase), which in a diploid cell would correspond to discriminating between a 1x an 1.5x tXA dose. Inactivation of exactly two out of four X chromosomes in tetraploid cells can be reliably ensured, if both, the 0.5x tXA and the 1.5x tXA doses lie outside the bistable regime in diploid cells (Fig

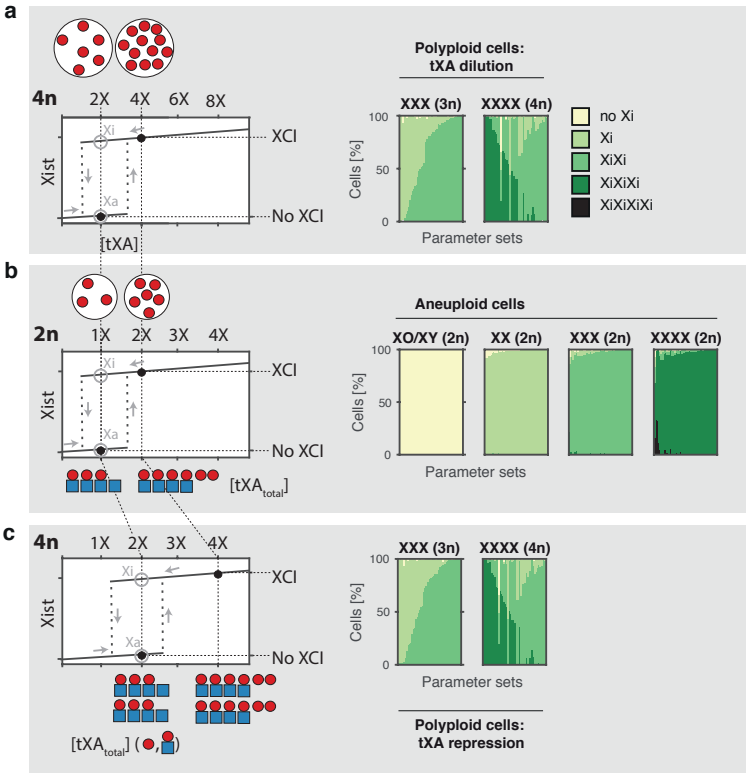

FIGURE 2.6: The cXR-tXA model can recapitulate Xist patterns in aneuploid and polyploid cell lines. a) Dilution hypothesis to explain XCI patterns in polyploid cells (see below). b) Simulation of diploid cells with 1, 2, 3 or 4 X chromosomes (male, female, X trisomy, X tetrasomy). Left: Schematic depiction of the bistable switch-like XCI response required to make an all-or-nothing decision on whether to initiate XCI between a single and a double tXA dose. Red circles: tXA; blue sqares: autosomal Xist repressor. Each X chromosome was assumed to produce an additional tXA dose, resulting in Xist upregulation until only a single active X (1x tXA) remains. Right: Simulation results. Each chromosome was classified as Xist+ or Xist-, the stacked bar graphs display the fraction of cells that fall into each category for 50 simulated parameter sets, that can generate reliable monoallelic Xist upregulation in diploid female cells. a,c) Dilution or molecular titration can change the total tXA stimulus range in which the system is bistable and can thereby explain XCI patterns in polyploid cells. Left: Schematic depiction of the alternative assumptions. a) assumes that the increased genome size results in an increased nuclear volume and therefore in tXA dilution, so that the effective tXA concentration in 4n4X cells and 2n2X cells is equal. c) assumes the existence of an autosomal Xist repressor (blue squares) that counteracts tXA, so that tXA is completely sequestered in both male diploid (2n1X) and tetraploid (4n2X) cells, while unsequestered tXA exists in female diploid (2n2X) and tetraploid (4n4X) cells. Right: Stacked bar graphs show the classification of Xist expression patterns in triploid (left) and tetraploid (right) cells for each of the assumed scenarios.

2.7b). In this case, the only stable cellular state is the XaXaXiXi state. In the XaXiXiXi state (=0.5x tXA in 2n), only the $Xist^{low}$ state is stable, resulting in Xist downregulation from the Xis. In the XaXaXaXi state (=1.5x tXA in 2n), only the $Xist^{high}$ state is stable, triggering further upregulation of Xist from the Xas. If, however, the bistable regime extends to a 1.5x tXA dose in diploid cells (Fig 2.7a, top), the XaXaXaXi state will also be stable in tetraploid cells, which can result in inactivation of too few X chromosomes (Fig 2.7a, bottom). Conversely, if the bistable regime extends to a 0.5x tXA dose in diploid cells (Fig 2.7c, top), the XaXiXiXi state in tetraploid cells is also stable, which can result in inactivation of too many X chromosomes (Fig 2.7c, bottom).

Thus, the model correctly reproduces polyploid XCI patterns in the majority of simulated sets. It also predicts that it is more challening to inactivate the correct number of X chromosomes in tetraploid than in diploid cells. In agreement with this, experimental data of differentiating

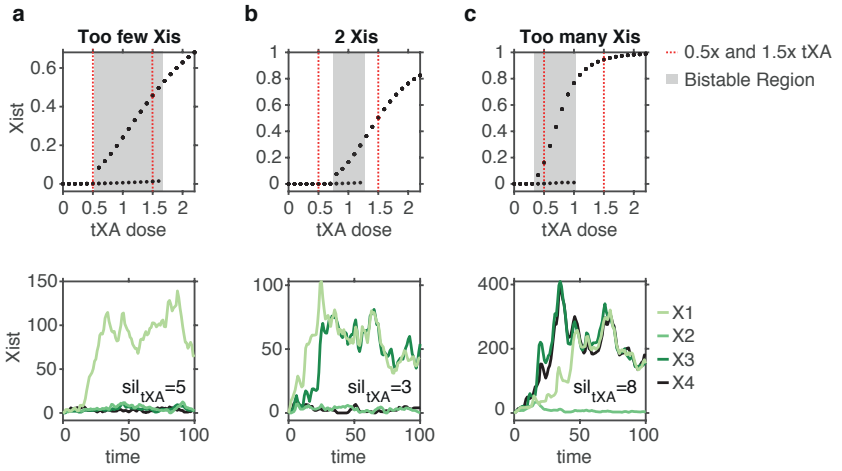

FIGURE 2.7: Inactivation of the correct number of X chromosomes is more challening in tetraploid than in diploid cells. Top: Allelic steady states in diploid cells simulated with the ODE model for a parameter set that correctly inactivates two (b), too few (a), or too many (c) X chromosomes in tetraploid cells. tXA dose corresponding to the number of active X chromosomes in diploid cells are indicated. Grey shading indicates bistable region, dotted red lines indicate steady states for 0.5x tXA (=1x tXA in 4n4X) and 1.5x tXA (=3x tXA in 4n4X). Bottom: Example Xist trajectories in an individual cell are shown. Different shadings of green represent the 4 Xist alleles.

tetraploid ESCs and tetraploid embryos report a considerable number of cells that inactivate too many or too few X chromosomes [107, 138].

2.4 cXR-tXA model recapitulates XCI patterns in different species

As discussed in section 2.2, the cXR-tXA model allows cells to reverse biallelic Xist expression through complete tXA silencing. Such transient biallelic Xist expression is reminiscent of the XCI pattern observed in rabbit embryos, where a majority of cells transiently express Xist from both X chromosomes and then resolve this biallelic to monoallelic expression. In fact, we found parameter sets that exhibited transient biallelic Xist upregulation to varying degrees but then reached the monoallelic expression state in a reliable fashion (Fig 2.8a, and Fig 2.8c for an example simulation). Possibly, our network could thus constitute the core framework of XCI regulation that is conserved across species. To test this tempting hypothesis we set out to better characterize the extent of transient biallelic Xist expression in our model as well as in mouse embryos.

In the model, the degree of transient biallelic Xist expression among the monoallelic parameter sets is determined by the relative time scales of Xist upregulation (switch-on) and tXA silencing (Fig 2.8b). Whether Xist is upregulated directly in a monoallelic fashion or transiently expressed biallelically depends on whether tXA silencing or upregulation of Xist from the second allele occur first. If Xist upregulation is much slower than tXA silencing (orange scenario in Fig 2.8b-c), the first allele to upregulate Xist will have reliably silenced tXA before the second allele can switch on Xist, resulting in direct monoallelic Xist upregulation. This however means that Xist upregulation must be slow. If Xist upregulation is faster, the second allele might also upregulate

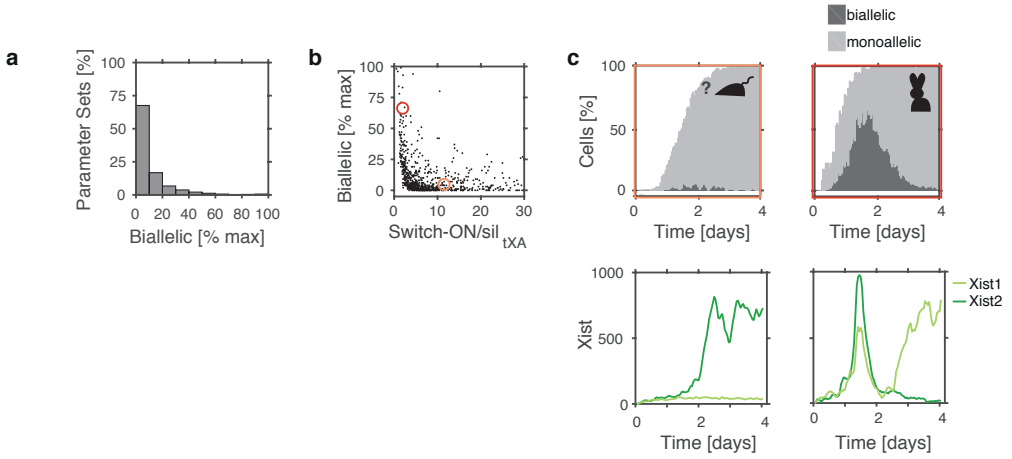

FIGURE 2.8: The cXR-tXA model can produce different extent of transient biallelic Xist expression. a) Distribution of the maximal percentage of cells with biallelic Xist expression observed during the simulation across all parameter sets, that correctly end up with monoallelic Xist expression in >80% of cells. b) For all monoallelic parameter sets, the maximal fraction of cells with biallelic Xist expression is plotted over the ratio of Xist switch-on time and tXA silencing delay (sil_{tXA}). The Xist switch-on time is defined as the first timepoint when Xist levels reach 20% of the $Xist^{high}$ state from the ODE simulation. Colored circles indicate the parameter sets for which example simulations are shown in c. c) Simulation of two example parameter sets depicting the fraction of monoallelic (light grey) and biallelic (dark grey) cells in a population of 100 simulated cells (top), or Xist trajectories in an individual cell (bottom). If Xist switch-on is slow and tXA silencing is relatively fast (= high switch-on/sil_{tXA}), Xist switch-on will usually occur on one allele at a time and the monoallelic expression state will be quickly stabilized by tXA silencing, resulting in a low frequency of biallelic Xist expression, as observed in mice (orange). If Xist upregulation is rapid, and silencing slower (=low switch-on/sil_{tXA}), Xist will frequently be expressed from both alleles initially, as observed in rabbits (red).

Xist before the first allele has silenced tXA. Thus, cells must compromise between speed and reliability of monoallelic Xist upregulation.

Possibly, the timescales of Xist upregulation and tXA silencing differ between species, meaning that they employ different strategies in this trade-off (compare red, orange in Fig 2.8b-c). Mouse embryos have been described to upregulate Xist in a direct monoallelic fashion (Fig 2.8b-c, orange) while in rabbit a majority of cells transiently upregulates Xist biallelically (Fig 2.8b-c, red). Interestingly, in the mouse in vitro model for random XCI, differentiating mESCs, transient biallelic Xist expression also occurs, and its extent depends on the precise culture conditions [139, 140]. This indicates that small changes in the regulation can also shift mouse cells to a regime with more biallelic expression, and that they might exhibit some low degree of transient biallelic Xist expression also in vivo which would be in agreement with a recent study [141]. It indeed seems plausible that the regulation would be tuned to a point where speed and accuracy are at an optimum, accepting that occasionally a cell will upregulate Xist from both alleles rather than making the regulation very reliable but extremely slow.

We therefore hypothesized that likely also in mouse embryos a fraction of cells transiently upregulate Xist biallelically. To investigate this, our collaborateurs in Japan[2] analyzed the Xist expression pattern in vivo in the E5.0 epiblast shortly after random XCI initiates. Indeed, RNA FISH revealed that 15-20% of cells in each embryo possessed two Xist clouds (Fig 2.9). Thus, transient biallelic Xist expression also occurs during the establishment of random XCI in mouse development but to a lesser extent than in rabbit or human, likely because the timescales of Xist

[2]Ikuhiro Okamoto and Mitinori Saitou

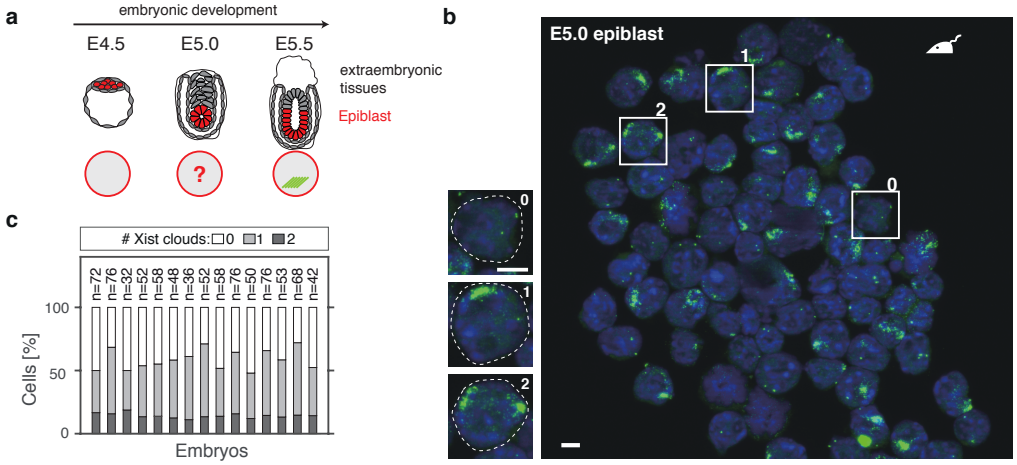

FIGURE 2.9: Biallelic Xist expression in E5.0 mouse embryos. a) In mouse development, monoallelic Xist upregulation occurs between E4.5-E5.5 in the cells of the epiblast (red). b) Xist/Tsix RNA FISH (green) and DAPI staining (blue) of female E5.0 epiblast cells. Scale bar, 5μm. White boxes indicate example cells with 0, 1 or 2 Xist clouds that are enlarged on the side, where dashed white lines mark the nucleus. c) Quantification of the percentage of cells with 0, 1 or 2 Xist clouds in the epiblasts of 15 female embryos. The number of cells that were counted in each embryo is given above (n).

upregulation and tXA silencing are different, i.e. because Xist upregulation is slower or tXA silencing is faster in mouse than in rabbit.

To investigate whether the cXR-tXA model could indeed reproduce quantitative expression data in different species, we compared the simulations to RNA FISH data from rabbit and mouse embryos [99, 142], and from differentiating mESCs. We compared the fraction of Xist mono- and biallelic cells between our simulations and experimentally measured timepoints. Since the developmental timepoint at which random XCI initiates is not determined precisely, we aligned simulations and experiments in time by allowing an offset of up to 24 h between the simulation and the experimental measurements (Fig 2.10a). Each parameter set was compared to the data with different offsets. To find the parameter set - offset combination that could explain the experimental data best we used a maximum likelihood estimate (MLE) approach (see section 9.2.4). Indeed, the model could reproduce all three experimental data sets by assuming different values for the reaction rates (Fig 2.10b).

We went on to test whether the model could also explain onset of random XCI in humans. Human embryos exhibit even more extreme biallelic Xist expression which persists over several days without inducing complete gene silencing [99, 100]. We do not know yet, how the biallelic XIST expression state is resolved to the XaXi state in humans, since culture conditions that allow us to observe this transition in vitro have only recently been established. In this new in vitro model, the transition from the biallelic XIST expression state to the XaXi state appears to occur via a state without XIST expression [101]. Our model can possibly help to identify alternative scenarios that could govern XCI regulation in humans. We tested two regulatory hypotheses to explain the extended biallelic XIST expression in human embryos: (1) cXR upregulation, where cXR only becomes upregulated at the time when biallelic XIST expression is resolved to monoallelic expression (Table 2.6 and Fig 2.11, left), and (2) cXR dampening where cXR is already expressed but also partially silenced by XIST during the biallelic period (Table 2.6 and Fig 2.11, right). Both scenarios assume that tXA is already expressed during the biallelic expression

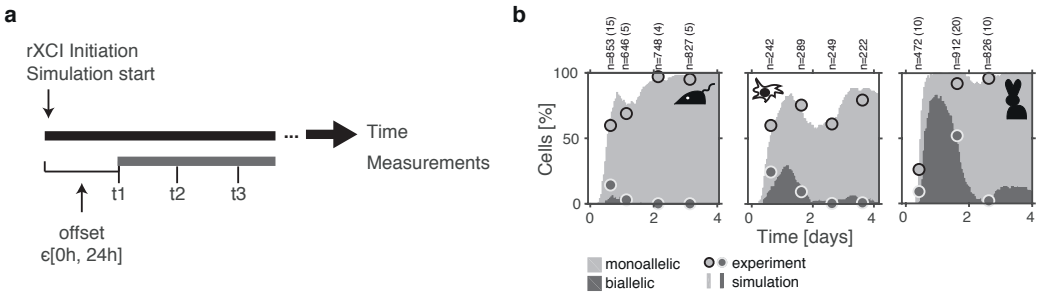

FIGURE 2.10: The cXR-tXA model can reproduce Xist expression data from different species. a) Scheme depicting the alignment of simulations and experimental data. An offset of up to 24 h was tested. b) Fraction of cells with mono- and biallelic Xist expression (light and dark grey, respectively) at different timepoints during mouse (left) and rabbit (right) development and during differentiation of mESCs (middle). The plots show the experimental data (circles) together with a simulation using the parameter set that best explains the data. The total number of cells analyzed for each timepoint is given above (n), for the in vivo data together with the number of embryos from which the cells were pooled (in parentheses). Experimental data from Fig 2.9, [99, 142], and differentiation of the Tx1072 line (three pooled independent replicates, see Fig 2.12).

TABLE 2.6: Equations for simulation of two alternative XCI scenarios in humans

	(1) cXR upregulation/activation	**(2) cXR dampening**
tXA	for $t < t_1$ $\frac{dtXA_i}{dt} = p_{23} - tXA_i$ for $t \geq t_1$ $\frac{dtXA_i}{dt} = p_{23}(1 - \frac{x_i^{p_3}}{x_i^{p_3} + (p_{21} \cdot p_4)^{p_3}}) - tXA_i$	for $t < t_1$ $\frac{dtXA_i}{dt} = p_{23} - tXA_i$ for $t \geq t_1$ $\frac{dtXA_i}{dt} = p_{23}(1 - \frac{x_i^{p_3}}{x_i^{p_3} + (p_{21} \cdot p_4)^{p_3}}) - tXA_i$
cXR	for $t < t_1$ $\frac{dcXR_i}{dt} = 0 - cXR_i$ for $t \geq t_1$ $\frac{dcXR_i}{dt} = p_{22}(1 - \frac{x_i^{p_5}}{x_i^{p_5} + (p_{21} \cdot p_6)^{p_5}}) - cXR_i$	for $t < t_1$ $\frac{dcXR_i}{dt} = p_{22}(1 - p_{34} \cdot \frac{x_i^{p_5}}{x_i^{p_5} + (p_{21} \cdot p_6)^{p_5}}) - cXR_i$ for $t \geq t_1$ $\frac{dcXR_i}{dt} = p_{22}(1 - \frac{x_i^{p_5}}{x_i^{p_5} + (p_{21} \cdot p_6)^{p_5}}) - cXR_i$

period, but not silenced or dampened yet. Table 2.6 summarizes the equations used to simulate the two alternative scenarios with the 100 parameter sets that could best reproduce monoallelic Xist upregulation in mice (see section 2.4). t_1 marks the timepoint at which complete silencing sets in and cXR is upregulated. In the dampening model, p_{34} describes the maximal reduction of cXR by XIST during the biallelic expression period. p_{34} was randomly drawn from a uniform distribution between 0.01 and 0.99 for each simulated parameter set.

Both scenarios can explain extended biallelic expression and the transition to the monoallelic expression state (Fig 2.11). Once tXA silencing sets in, biallelic XIST expression is quickly destabilized and stable monoallelic XIST expression and cXR/tXA silencing is established. For the two alternative models, Fig 2.11 shows XIST trajectories in an individual cell (top), the fraction of mono- and biallelic cells in a population of 100 cells with the same parameter set (middle), and the percentage of mono- and biallelically XIST expressing cells in all 100 simulated parameter sets (bottom).

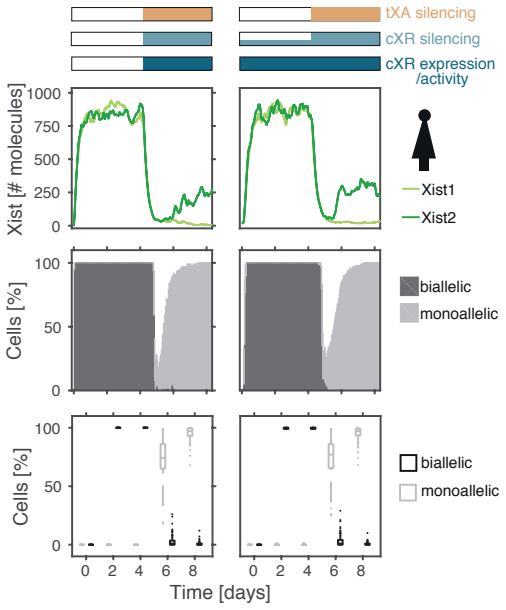

FIGURE 2.11: Simulation of two alternative scenarios that can explain persistent biallelic Xist expression in human embryos during the dampening period, where Xist-mediated silencing is limited. Left: cXR only starts to be expressed once complete silencing sets in. Right: In the initial period of biallelic Xist expression cXR is partially silenced (dampened) by Xist, while tXA is unaffected. In both scenarios the biallelic expression is resolved to monoallelic expression once complete silencing (of tXA and cXR) sets in. Simulations of an individual cell (top) and a population of 100 cells (middle) are shown for one example parameter set. The bottom boxplots show the percentage of mono- and biallelically Xist expressing cells in all 100 simulated parameter sets. In each box, the center line represents the median; boxes extend to upper and lower quartiles; whiskers indicate the most extreme points not considered outliers; points represent outliers.

In conclusion, the cXR-tXA model can reproduce different degrees of biallelic Xist expression observed across different mammalian species and therefore provides a unifying framework of XCI regulation. We propose two alternative scenarios on how biallelic Xist expression might be resolved to monoallelic expression in human embryos. Both of these scenarios are in agreement with the recent observation that hESCs seem to transit from the XaXa state with biallelic XIST expression to the XaXi state with monoallelic XIST expression via a state without XIST expression [101].

2.5 Experimental testing of model predictions

2.5.1 Xist switch-on time controls extent of biallelic expression

An essential prediction of the cXR-tXA model is that the extent of biallelic Xist expression can be modulated by the switch-on time of Xist, or more precisely the time period between switch-on of the first and second allele. The longer this period, the less frequently biallelic Xist upregulation is expected to occur (given that the kinetics with which tXA is silenced remain the same). Here we set out to validate the effect of the Xist switch-on time on the extent of biallelic Xist upregulation experimentally by artificially modulating the switch-on-to-tXA-silencing ratio in the following way[3]: We used a system that allows us to accelerate Xist upregulation from one allele, as this will prolong the time before Xist is upregulated from the other allele, which is predicted to result in an increased switch-on-to-tXA-silencing ratio and a lower extent of biallelic expression (see Fig 2.8b in section 2.4). We used the Tx1072 mESC line that is a hybrid between two polymorphic strains (C57BL6/J X Cast/EiJ), which will be called B6 and Cast here. On the B6 X chromosome this cell line carries a doxycycline-inducible *Xist* promoter, such that Xist upregulation can be accelerated in *cis* by doxycycline treatment (Fig 2.12a). When differentiated from 2i medium, where the cells are in a ground pluripotent state, the Tx1072 cell line recapitulates random XCI with quite a high fraction of cells passing through a transient biallelic Xist expression state (Fig 2.12c, black). Doxycycline induction artificially decreases the Xist switch-on time from B6, which will consequently prolong the period between switch-on of the first (B6) and the second (Cast) allele (and thereby allows us to test the impact of the switch-on time on the extent of biallelic expression) (experimental setup in Fig 2.12a).

Table 2.7: Equations for simulation of doxycycline induction

	Chr X_1 (B6)	**Chr X_2 (Cast)**
Xist	$\frac{dx_1}{dt} = p_{21} \cdot (p_{24} + p_{29} \cdot f(cXR_1) \cdot f(tXA)) - x_1$	$\frac{dx_2}{dt} = p_{21} \cdot (p_{30} \cdot f(cXR_2) \cdot f(tXA)) - x_2$

We simulated the experimental setup with our cXR-tXA model using the 100 parameter sets that best fit the mESC data (section 2.4). To this end, we assumed that doxycycline induction will increase the Xist initiation rate on the B6 allele. The model was modified as summarized in Table 2.7.

Table 2.8: Parameter values for simulating doxycycline induction

	+ Dox		**Control**	
	t<0	**t≥0**	**t<0**	**t≥0**
p_{24}	10	10	0	0
p_{29}	0	0	0	1
p_{30}	0	1	0	1

p_{24} determines doxycycline regulation of Xist on the B6 and is set to 10 in the presence of doxycycline and to 0 in the absence of doxycycline. p_{29} and p_{30} control the regulation of Xist by cXR

[3]The doxycycline induction experiment was performed by Edda Schulz.

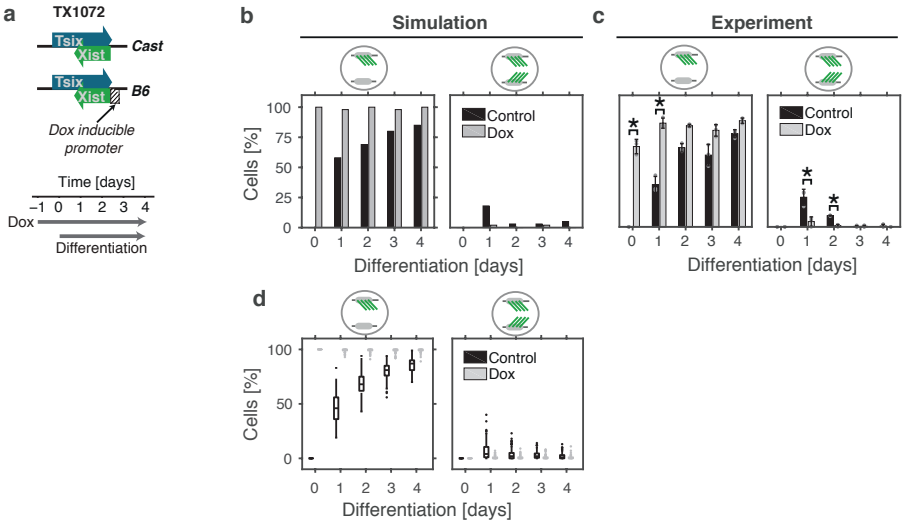

FIGURE 2.12: Premature Xist upregulation by doxycycline treatment reduces the extent of biallelic Xist expression. a) Scheme of the cell line used (top) and the experiment performed (bottom). The Tx1072 cell line is a Cast x B6 hybrid and carries a doxycycline-inducible *Xist* promoter on its B6 allele. One day prior to differentiation Xist from the B6 allele was induced with doxycycline. b) Simulation of the experimental setup with one example parameter set. c) Experiment conducted as described in a). Xist expression was assessed by RNA FISH. Mean and standard deviation of three independent experiments are shown. *: $p<0.05$ in two-sample two-sided t-test. b-c) The percentage of monoallelically (left) and biallelically (right) Xist expressing cells are shown for the uninduced control (black) and the dox induced (grey) population. d) Simulation results for all 100 simulated parameter sets. The boxplots show the percentage of monoallelically (left) and biallelically (right) Xist expressing cells in doxycycline-treated (grey) and untreated control (black). In each box, the center line represents the median; boxes extend to upper and lower quartiles; whiskers indicate the most extreme points not considered outliers; points represent outliers.

and tXA on B6 and Cast, respectively, and are set to 1 on an allele that is not induced with doxycycline, and to 0 on an allele that is induced with doxycycline, because upon doxycycline induction, Xist expression likely becomes independent of tXA and cXR. Before differentiation is induced p_{29} and p_{30} are also set to 0, representing the action of pluripotency factors that repress Xist in undifferentiated cells (Table 2.8).

The model predicts that upon doxycycline induction the extent of biallelic expression will decrease compared to the uninduced cells, because Xist on B6 has more time to silence tXA before Cast will also upregulate Xist (Fig 2.12b for one example set and Fig 2.12d for all 100 simulated sets). Indeed, when cells were induced with doxycycline 1 day prior to induction of differentiation, the fraction of cells that were captured in a biallelic expression state was reduced from ∼25% to less than 5% of cells (Fig 2.12c). This confirms that the speed of Xist upregulation modulates the extent of biallelic Xist expression.

2.5.2 Biallelic Xist upregulation is reversible

The model predicts that cells which accidentally upregulated Xist from both alleles are able to revert this biallelic expression state by complete tXA silencing. In fact, the fate (and even the existence) of these biallelic Xist expressing cells has been a matter of extensive debate in the field of XCI [99, 140]. A stochastic XCI regulation will result in a fraction of cells with too many or too few Xist expressing alleles. If cells are not able to correct such erroneous Xist expression

patterns, a secondary mechanism would need to be in place to counter-select these cells. The significant size difference between female and male early post-implantation embryos at a developmental timepoint before hormone production starts to influence growth [143–145], has been interpreted as evidence for such cell loss caused by selection during the XCI process [107, 140]. The size difference between the sexes can however also be explained by the halt of differentiation in female embryos until proper dosage compensation has been achieved which slows down female embryogenesis early after implantation [58, 143, 146, 147]. Others have argued that the invoked counter selection against cells with erroneous Xist expression patterns would be too costly for the embryo and have used this argument to eliminate a stochastic model for XCI regulation [148]. In this context, our model prediction that cells are able to revert biallelic to monoallelic Xist expression is of great interest, as it could solve this dilemma.

To test this prediction we employed a system that allowed us to artificially increase the extent of biallelic expression and to then investigate the fate of these cells. To this end, we used the Tx1072dT cell line which was derived from the Tx1072 line by deletion of the DxPas34 region from the Cast allele (Fig 2.13a). DxPas34 acts as an enhancer for Tsix and its heterozygous deletion has been shown to decrease Tsix transcription and result in preferential Xist upregulation from the mutant X chromosome [20]. On the B6 allele this cell line - just as its parental line - carries a doxycycline-inducible *Xist* promoter. We differentiated the cells for 48 h, triggering Xist upregulation from the Cast allele.[4] Then we added doxycycline to also induce Xist expression from the B6 allele (Fig 2.13b). This increased the fraction of biallelic cells from $\sim$ 12% to 30% 6 h after doxycycline addition (Fig 2.13c).

To reproduce this experimental setup in our simulations, we set p_{29} to 33% to reduce Xist transcription in *cis* on the B6 allele, thereby recapitulating the initial skewing of Xist upregulation due to the heterozygous *Tsix* deletion on the Cast chromosome. Table 2.9 summarizes the parameter values that were used for the simulation. We simulated the experiment with the 100 parameter sets that best fit the mESC data (section 2.4).

Table 2.9: Parameter values for simulating forced biallelic Xist expression

		t<48h	**t≥48h**
Xist induction by doxycycline on B6	p_{24}	0	10
Xist regulation by cXR and tXA on B6	p_{29}	0.33	0
Xist regulation by cXR and tXA on Cast	p_{30}	1	1

The model predicts that since Xist expression on the B6 is artificially maintained by doxycycline, cells will downregulate Xist from the Cast allele to resolve the biallelic expression state (Fig 2.13d left, e). We used amplicon sequencing of single-nucleotide polymorphisms (SNPs) on cDNA to quantify Xist expression in an allele-specific manner. The assay revealed that Xist was significantly downregulated from the Cast allele compared to the untreated control 48 h after the start of doxycycline induction in agreement with our model prediction (Fig 2.13d, right).

To provide more direct evidence that cells had indeed downregulated Xist from the Cast allele rather than being counter-selected and gradually lost from the population as a result of having inactivated both X chromosomes, we performed RNA FISH for Xist together with immunostaining for H3K27me3. H3K27me3 is a characteristic repressive mark that is recruited

[4] The experimental data depicted in Fig 2.13 was generated by Ilona Dunkel.

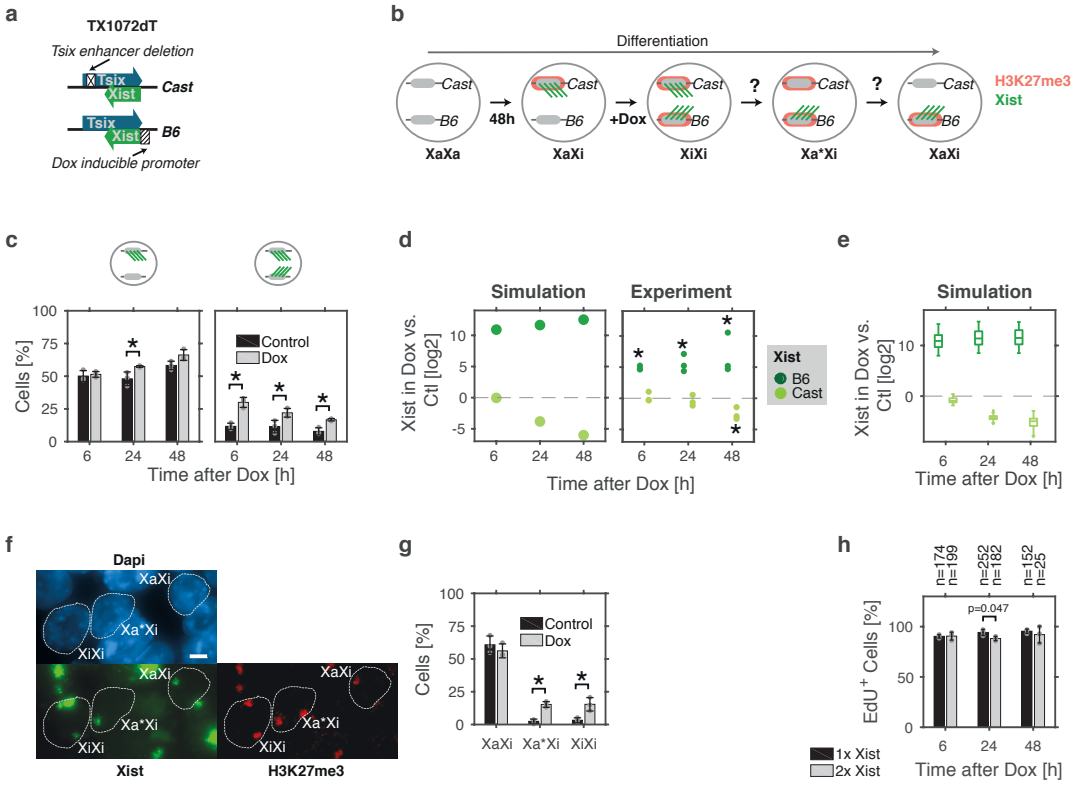

FIGURE 2.13: Biallelic Xist upregulation is reversible. a-b) The cell line and experimental setup used to artificially induce biallelic Xist upregulation and investigate the fate of biallelically Xist expressing cells. a) The Tx1072dT line is derived from the Tx1072 by deletion of the Tsix enhancer *DxPas34* from the Cast allele. b) Tx1072dT cells were differentiated for 48 h, resulting in Xist (green) upregulation from the Cast allele, and recruitment of H3K27me3 (red) to the future Xi. After 48 h of differentiation, cells are treated with doxycycline to force Xist upregulation also from the B6. The model predicts that this biallelic expression state is resolved by Xist downregulation from the Cast allele. Cells might transit through a state where H3K27me3 is still enriched on both X chromosomes, but Xist has already been downregulated from the Cast allele (Xa*Xi). c) Xist expression patterns at different timepoints after the start of the doxycycline treatment measured by RNA FISH. The percentage of monoallelically (left) and biallelically (right) Xist expressing cells in the control (black) and the doxycycline-treated (grey) population. Mean and standard deviation of three independent experiments are shown (>100 cells per replicate). *:p<0.05 in two-sample two-sided T-test. d) Simulation (left) and experimentally measured (right) levels of Xist expression from the B6 and Cast alleles at different timepoints after doxycycline addition. To determine allele-specific Xist expression experimentally, allele-specific amplicon sequencing was used. *:p<0.05 in one-sample two-sided T-test. For the simulation, one example parameter set is shown in d), while the boxplot in e) shows the simulation results for all 100 tested parameter sets. In each box, the center line represents the median; boxes extend to upper and lower quartiles; whiskers indicate the most extreme points not considered outliers; points represent outliers. f-g) Immunofluorescence for H3K27me3 together with Xist RNA FISH 48 h after the addition of doxycycline. f) Example image showing a cell of each classification category. XiXi: Two Xist clouds, two H3K27me3 signals; Xa*Xi: One Xist cloud, two H3K27me3 signals; XaXi: One Xist cloud, one H3K27me3 signal. Scale bar, 5μm. g) Quantification of the XiXi, Xa*Xi and XaXi states. Mean and standard deviation of three independent experiments are shown (>120 cells per replicate). *:p<0.05 in two-sample two-sided T-test. h) EdU treatment of cells to assess their proliferation, by measuring EdU incorporation into the DNA during replication. The barplot shows the quantification of the percentage of EdU positive cells at the indicated timepoints in cells expressing Xist monoallelically (black) and biallelically (grey). Mean and standard deviation of three independent replicates are shown (>50 cells per replicate, except at 48h), paired two-sample two-sided T-test.

to the future inactive X chromosome by the Xist RNA. In the RNA FISH we also included a probe for Huwe1 which is a late-silenced X-linked gene and can therefore be used to mark the X chromosome. We hoped to be able to capture cells that had already downregulated Xist from the Cast allele, but still retained the histone mark on both X chromosomes. This would allow us to conclude that these cells in the past had also expressed Xist from the second chromosome because it retains the H3K27me3 mark, but do not any more because Xist RNA is gone. We indeed identified such cells (Xa*Xi, Fig 2.13f), and they occurred much more frequently in the sample that was treated with doxycycline to induce biallelic Xist expression, than in the control cells that were differentiated without doxycycline addition (Fig 2.13g). Interestingly, 6 h after doxycycline induction, ∼30% of cells expressed Xist biallelically (Fig 2.13c), while 48 h after doxycycline addition half of these previously biallelic cells (15% of the total population) had resolved to monoallelic expression (Xa*Xi), while ∼15-20% still expressed Xist biallelically (Fig 2.13c,g).

As an additional proof against selection as a major reason for the depletion of biallelic cells from the population over time, we assessed cell replication capacity via 5-ethynyl-2'-deoxyuridine (EdU) incorporation during replication, and found that among the biallelic cells only a slightly lower fraction was EdU positive than among the monoallelic cells 24 h after doxycycline induction (88% vs 94%, Fig 2.13h), indicating that slower proliferation or cell death of biallelic cells cannot explain the loss of these cells from the population.

In summary, we have thus shown that the speed of Xist upregulation indeed modulates the extent of biallelic Xist expression. We have also provided the first direct experimental evidence for the reversal of biallelic to monoallelic Xist expression, as a mechanism to cope with occasionally occurring erroneous Xist expression patterns that arise from stochastic Xist upregulation.

2.6 Tsix as a cXR candidate

Our approach has uncovered the minimally required regulator types for monoallelic Xist upregulation. To rigorously test our proposed cXR-tXA model, we now have to identify the molecular factors that act as the predicted regulators. For the tXA, no gene with all predicted characteristics has been identified yet (see discussion 3.3 for details). However, at least in mice, a good candidate for the *cis* repressor could be Xist's antisense transcript Tsix (Fig 2.14). Antisense repression between Xist and Tsix could generate the predicted double negative feedback required for bistability. Tsix transcription has been shown to establish a repressed chromatin state at the *Xist* promoter [25, 27, 28]. Xist RNA does silence Tsix in *cis*. Consistent with this hypothesis, Xist and Tsix show opposing expression patterns on Xa and Xi shortly after XCI has initiated with Tsix being expressed from the Xa and Xist being expressed from the Xi. The exact molecular details of repression between Xist and Tsix are however still unknown.

To understand whether Tsix could indeed mediate the predicted double negative feedback we developed a detailed mathematical model of the *Xist/Tsix* locus describing RNA polymerase II (RNAP) initiation, elongation and RNA degradation of both genes. We assumed three different mechanisms of repression: (1) Xist RNA-mediated silencing of the *Tsix* promoter, (2) Tsix

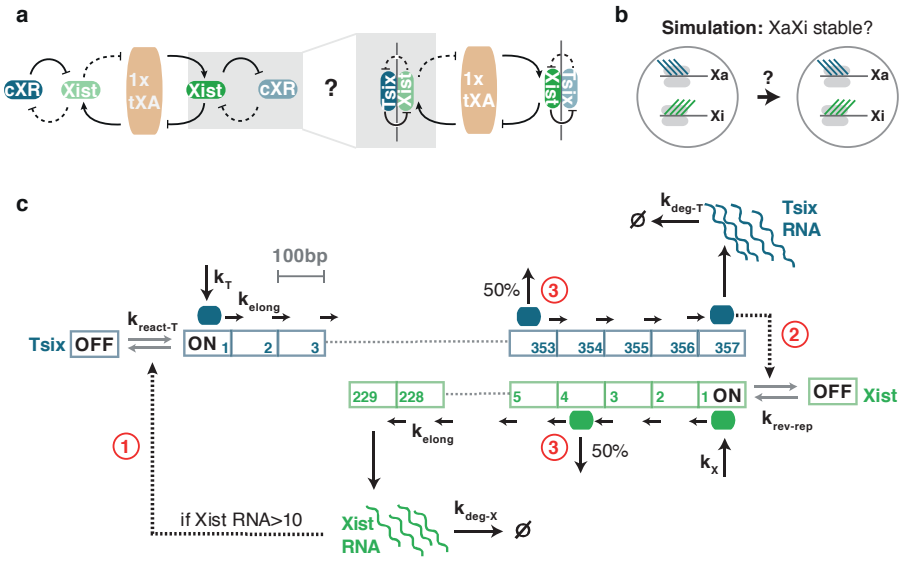

FIGURE 2.14: Antisense transcription as the mediator of the local double negative feedback. a) Schematic representation of the cXR-tXA network. Potentially, Xist's antisense transcript Tsix can act as cXR. b) Scheme of the simulation to analyze whether mutual repression between Xist (green) and Tsix (blue) can stably maintain an Xa with high Tsix and low Xist expression, and an Xi with high Xist and low Tsix expression. c) Scheme of the discrete stochastic model to simulate Xist and Tsix transcription. RNAPs can bind to both promoters (k_T, k_X), elongate along the gene (k_{elong}), and produce a full-length RNA when they reach the end of the gene. RNAs get degraded with first-order reaction kinetics (k_{deg-X}, k_{deg-T}). Both promoters exist in a transcriptionally active (ON) and inactive (OFF) state. Xist and Tsix repress each other by three mechanisms: The Xist RNA induces transition of the *Tsix* promoter to the OFF state (1), passing Tsix RNAPs turn the *Xist* promoter OFF (2). Collision of two antisense RNAPs results in dislodgement of one randomly chosen RNAP (3). Promoter OFF states are reverted to the ON state in the absence of Xist RNA or Tsix RNAPs with rates $k_{react-T}$ and $k_{rev-rep}$.

transcription-mediated repression of the *Xist* promoter and (3) transcriptional interference (TI) by RNAP collisions (Fig 2.14c). While the RNAP elongation rate and the degradation rates of Xist and Tsix RNA were previously estimated experimentally [85, 134, 149], the Xist and Tsix transcription initiation rates are unknown and were varied between 10 - 1000 RNAPs h^{-1}. Both promoters exist in a transcriptionally inactive OFF state where no transcription initiation can occur and an ON state in which RNAPs can bind to the promoter with the respective initiation rate [150, 151]. The *Xist* promoter transits to the OFF state when passed by Tsix antisense RNAPs, the *Tsix* promoter is turned off by Xist RNA-mediated silencing (Fig 2.14c). To modulate the strength of Tsix induced repression of the *Xist* promoter we varied the half-life of the OFF state between a few seconds to a few hours. *Tsix* promoter silencing by Xist can also be reverted with rate $k_{react-T}$ upon degradation of Xist RNA below a threshold of 10 molecules. Since Tsix overlaps the entire Xist gene of about 23 kb, antisense RNAPs will inevitably encounter each other on the DNA template (when the two promoters are simultaneously active). It has been observed experimentally that collision between antisense RNAPs can result in transcription termination [152–154]. We therefore included removal of one randomly chosen RNAP from the DNA template upon a collision as an additional previously uncharacterized layer of mutual repression between Xist and Tsix. Table 2.10 summarizes all parameters of the Xist/Tsix model with their respective ranges.

The reactions describing transcription initiation, transcription elongation and RNA degradation of Xist and Tsix were combined into a mathematical model to simulate individual cells. In the model, each gene was devided into segments of 100bp and elongation occurred discontinuously

Table 2.10: Parameters of the Xist/Tsix model for the XaXi simulation

Description	Parameter	Parameter Values
Xist Transcription rate [h^{-1}]	k_X	5, 6.35, 8.1, 10.35, 13.2, 16.8, 21.4, 27.3, 34.75, 44.3, 56.45, 71.9, 91.65, 116.8, 148.8, 189.65, 241.65, 307.95, 392.4, 500
Tsix Transcription rate [h^{-1}]	k_T	as k_X
Xist degradation rate [h^{-1}]	k_{deg-X}	0.1733 [85, 134]
Tsix degradation rate [h^{-1}]	k_{deg-T}	1.3868 [134]
Elongation rate [bp/s]	k_{elong}	40 [149]
Reversal rate of *Xist* promoter repression [h^{-1}]	$k_{rev-rep}$	0.1, 0.1438, 0.2069, 0.2976, 0.4281, 0.6159, 0.8859, 1.2743, 1.8330, 2.6367, 3.7927, 5.4556, 7.8476, 11.288, 16.238, 23.357, 33.598, 48.329, 69.519, 100

with RNAPs moving from one segment to the next at fixed time intervals of 2.5 seconds inferred from measurements of polymerase speed (elongation of one 100bp interval at $k_{elong} = 40$ bp/sec, Fig 2.14c). Between the elongation events, all other reactions were simulated using the stochastic Gillespie algorithm. In brief, RNAPs can bind to the promoter, elongate along the gene whilst interfering with antisense RNAPs by collisions, and at the end of the gene produce a full length RNA, that is degraded with a constant rate. Xist RNA represses the *Tsix* promoter. Tsix RNAPs switch the *Xist* promoter to the inactive state when passing it (Fig 2.14c).

2.6.1 Mutual repression between Xist and Tsix can maintain the monoallelic expression state

In a first step, we investigated whether mutual repression between Xist and Tsix can maintain the monoallelic Xist expression state (XaXi maintenance). We simulated the model with 8000 parameter sets, testing a range of values for the initiation rates and the strength of *Xist* promoter repression by Tsix RNAPs (Table 2.10). Since we simulated post-XCI cells where tXA silencing on the Xi has occurred, we assumed a constant single tXA dose. For each parameter set we simulated 100 cells for 500 h, with one chromosome starting from the Xa state, where Tsix is transcribed and Xist is repressed, and the second chromosome starting from the Xi state, where Xist is expressed and Tsix is silenced. A parameter set was classified as monoallelic if ≥99% of Xa/Xi pairs stably maintained the XaXi state (>10 Xist RNA molecules on Xi and <10 Xist RNA molecules on Xa). Approximately half of all tested parameter sets indeed maintained the alternative Xist expression states on Xa and Xi stably over the complete simulation (Fig 2.15a, example simulations of a stable and unstable parameter set in Fig 2.15b).

We then analyzed the parameter values that allow stable XaXi maintenance and found that the ratio of Xist and Tsix initiation rates (k_X/k_T) had to be smaller than 1 to maintain the XaXi state (Fig 2.15c, black). If $k_X > k_T$, Xist was upregulated from the Xa (Fig 2.15b,c, grey). Fig 2.15d plots Xist expression at the end of the simulation on the chromosomes that initiated as Xa (light green) or Xi (dark green), for different values of k_X, while k_T and $k_{rev-rep}$ were kept constant (k_T=113 h^{-1}, $k_{rev-rep}$=1 h^{-1}). It shows that if k_X is increased above a certain threshold the system switches from a regime where both, the $Xist^{high}$ state on Xa and the $Xist^{low}$ state on Xi, are stably maintained (grey) to a regime where the Xa becomes unstable

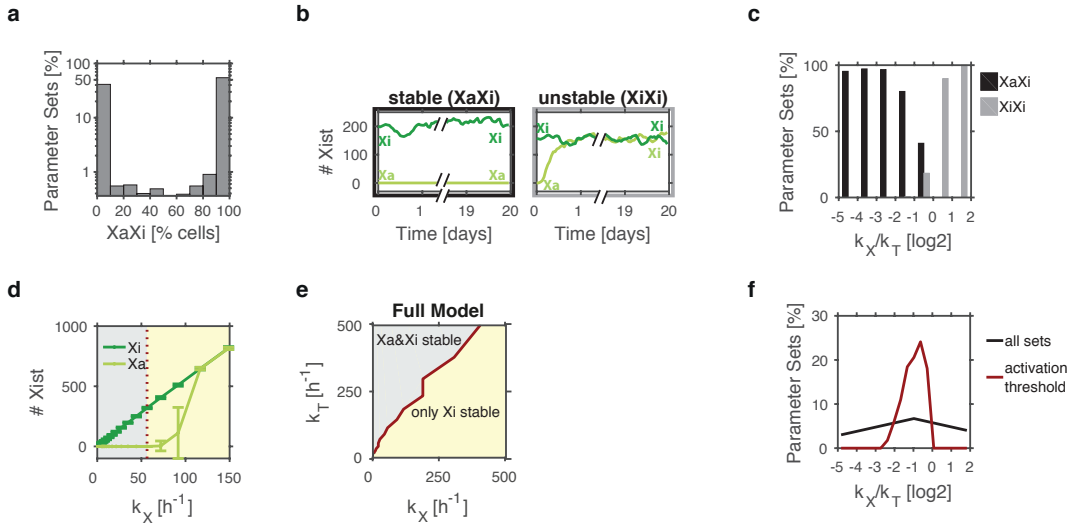

FIGURE 2.15: Mutual repression between Xist and Tsix can maintain the XaXi state. a) Distribution of the percentage of cells that stably maintain the XaXi state across all parameter sets. b) Two stochastic example simulations of Xist expression from Xa (light green) and Xi (dark green) in individual cells showing one set that stably maintains the XaXi state (black), and one set in which the Xa is destabilized (grey). c) Percentage of parameter sets that can stably maintain the XaXi state (black), and destabilize the Xa state (grey) for different k_X-to-k_T ratios. d) Mean and standard deviation of Xist expression from the X that initiated as Xa (light green) and the X that initiated as Xi (dark green) is shown for one example parameter set (k_T = 113 h^{-1}, $k_{rev-rep}$=1 h^{-1}) for different k_X values. The red dotted line indicates the k_X threshold above which the Xa state becomes unstable in >1% of simulated cells, and separates the regime where XaXi can stably be maintained (grey) from the regime where only the Xi is stable (yellow). e) The regime of stable XaXi maintenance (grey), and of Xa destabilization (yellow) plotted in the k_X-k_T space with the activation threshold (red line) as defined in d). f) Distribution of the k_X-to-k_T ratio at the activation threshold as defined in d) across all parameter sets.

and upregulates Xist (yellow). We defined an activation threshold of the k_X-to-k_T ratio above which at least 1% of cells upregulate Xist from the Xa (dotted red line in Fig 2.15d). Plotted in the k_X/k_T space, the activation threshold separates the regime where Xa and Xi are stable (grey) from the regime where only the Xi is stable (yellow, Fig 2.15e). Fig 2.15f plots the distribution of this activation threshold across all parameter sets showing that it lies at or below k_X-to-k_T ~1. Thus, if the k_X/k_T ratio is increased (by decreasing k_T or increasing k_X) a parameter set can shift from a regime of stable XaXi maintenance, to a regime where Xist upregulation is triggered from the Xa. A tXA-mediated change in the k_X/k_T ratio could thus be the mechanism that ensures that Xist upregulation is triggered in a female cell with two active X chromosomes, while the XaXi state can be stably maintained in a post-XCI cell.

Next, we investigated whether all repressive mechanisms that we had included were necessary to stabilize the XaXi state. To this end, we constructed six reduced model versions that contain only two, or a single repressive mechanism. Each of these reduced models was simulated with the same parameter values as the full model. The fraction of stable sets in each of the model versions is plotted in Fig 2.16a. Interestingly, we found that Xist-mediated silencing of the *Tsix* promoter was strictly required for XaXi maintenance, but was not sufficient (as it is missing the second arm of the double negative feedback). Thus at least two of the three mechanisms must be combined, and one of them must be Xist-mediated silencing of Tsix. One possible reason for the requirement of Xist RNA-mediated silencing could be that it is the only mechanism that is mediated by an RNA rather than by the act of transcription, and therefore can filter out fluctuations on the RNAP level that occur on faster timescales.

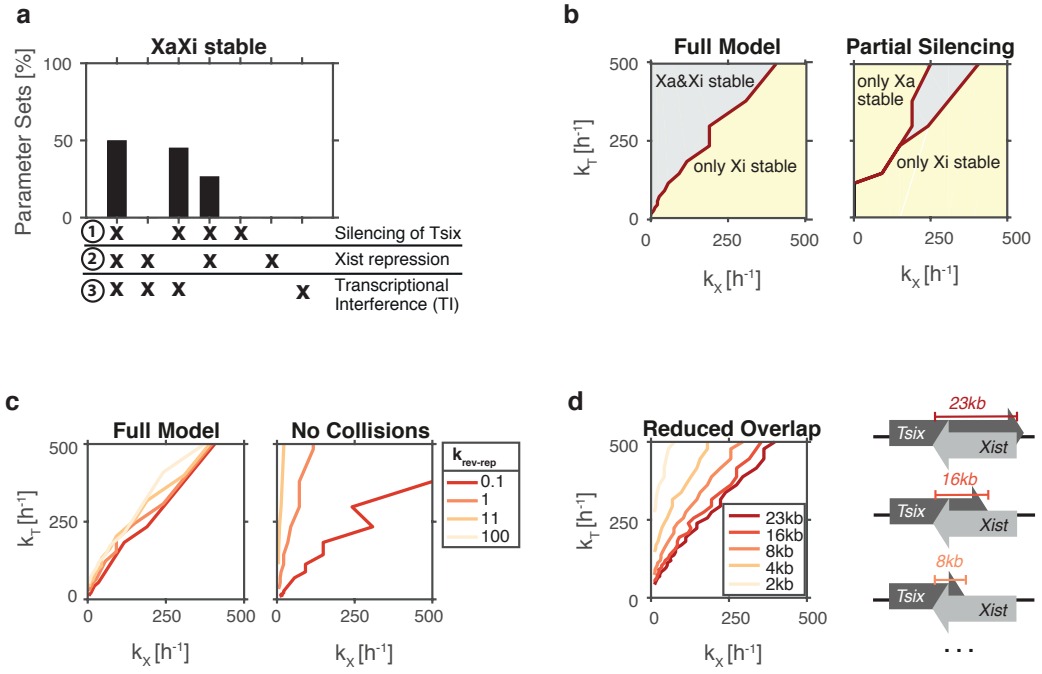

FIGURE 2.16: Reduced model versions. a) Percentage of parameter sets that maintain the XaXi state in >99% cells for the model with all repressive mechanisms and all possible model reductions with only two or one mechanisms of repression between Xist and Tsix. b) The regime of stable XaXi maintenance (grey), and of Xa or Xi destabilization (yellow) plotted in the the k_X-k_T space for the model with full (left) and partial (right) Tsix silencing. c) Effect of variation of the stability of the repressed *Xist* promoter OFF state in the full model (left) and the model without polymerase collisions (right). d) Effect of decreasing the overlap between Xist and Tsix in the model without *Xist* promoter repression. Colored lines in c-d indicate the threshold at which the Xa state becomes unstable in >1% of cells (for the respective $k_{rev-rep}$ value or overlap length).

We then investigated how a gradual reduction in the strength of the different repressive mechanisms would affect the ability of the system to maintain Xa and Xi stably. In humans, XIST has been reported to initially only dampen rather than silence X-linked gene expression. We therefore set out to understand whether partial silencing of Tsix by Xist RNA is sufficient to generate bistability. This could possibly explain the stable maintenance of the biallelic Xist expression state in humans, even in the presence of only a single tXA dose, which might result from partial silencing of tXA on both X chromosomes. Instead of inducing a transition of the *Tsix* promoter to the inactive OFF state, Xist RNA is now assumed to reduce the Tsix initiation rate by a certain fraction, given by the model parameter $sil_{partial}$. Although our previous analysis of the reduced model versions revealed that Xist RNA-mediated silencing is required for XaXi maintenance, Xist does not need to completely silence Tsix. Also, if Tsix is only partially silenced, a bistable (XaXi maintenance) region exists (Fig 2.16b right, grey) but it is smaller. In addition, for low Xist initiation rates a regime where only the Xa is stable appears. In contrast, reducing the stability of *Xist* promoter repression by Tsix, has no effect on the bistable region as long as Xist RNA-mediated silencing and polymerase collisions are present (Fig 2.16c, compare left to right), even if the repression strength ($k_{rev-rep}$) is varied over several orders of magnitude (Fig 2.16c, left).

It is still unclear whether Tsix is conserved in other mammals. In rabbits, no Tsix gene overlapping Xist has been identified so far. In humans, Tsix transcription has only been detected in embryonic carcinoma cells [29] where it has a different locus architecture compared to mouse: The overlap between XIST and TSIX is reduced and TSIX does not transcribe through the *XIST* promoter. Possibly, other regulators take over the role of Tsix in XCI in other species. One promising candidate could for instance be XACT in humans, which represses XIST in *cis* and is expressed in a mutually exclusive manner with XIST in post-XCI cells (XACT coats the Xa, XIST coats the Xi) [30, 31]. Nevertheless, we wanted to understand whether the reduced overlap that has been described for TSIX in human embryonic carcinoma cells, in principle is also capable of generating bistable XIST expression states on Xa and Xi. In our model, at least in the presence of polymerase collisions, repression of the *Xist* promoter by Tsix transcription was not required for bistability. We therefore simulated various alternative locus architectures with shorter overlaps where Tsix transcription is terminated before the *Xist* promoter and plotted the activation threshold at which at least 1% of cells had upregulated Xist from the Xa at the end of the simulation (Fig 2.16d). Indeed, a much shorter overlapping region can also generate bistability (2-4 kb are sufficient). However, the shorter the overlap, the higher must the Tsix initiation rate be relative to the Xist initiation rate to guarantee bistability (Fig 2.16d).

In summary, two repressive mechanisms (silencing combined with either collisions or promoter repression) are required for XaXi maintenance, but silencing does not need to be complete and the overlap between Xist and Tsix can be much smaller than in mouse. In addition, a decrease of the k_X-to-k_T ratio can shift the cell from a regime where Xist upregulation is triggered to a regime of stable XaXi maintenance, which is necessary to mediate the XaXa $\rightarrow$ XaXi transition. In the next section we include the tXA regulation, to see whether a double tXA dose can indeed trigger Xist upregulation in the XaXa state, and whether tXA silencing is then able to shift the cell from the regime where Xist upregulation is triggered (XaXa) to the regime where monoallelic expression is stable (XaXi), by decreasing the k_X-to-k_T ratio.

2.6.2 The Xist/Tsix model can reproduce monoallelic Xist upregulation

Just as for the general cXR-tXA model, we now sought to understand whether the Xist/Tsix model could reproduce monoallelic Xist upregulation (XaXa $\rightarrow$ XaXi). To couple the two X chromosomes, we included the negative feedback mediated by tXA silencing (Fig 2.17a). This was implemented by making the Xist initiation rate dependent on the tXA dose:

$$k_X^{eff} = k_X \cdot (tXA_1 + tXA_2)$$

The tXA concentration is modeled as a step function with value 1 if the respective allele is active and value 0 if it has been silenced by Xist, neglecting stochastic fluctuations in tXA production. This is justified either if tXA is a highly expressed gene, or if multiple tXA factors act together in an additive manner to activate Xist, such that random fluctuations in expression levels average away and robust dosage sensing within a two-fold range can be ensured. In the initial XaXa state, we set $tXA_{tot} = tXA_1 + tXA_2 = 2$. To incorporate repression of Xist by pluripotency factors, we set the Xist initiation rate in the undifferentiated state to $k_X^{eff} = 0.1 \cdot (tXA_1 + tXA_2) \cdot k_X$

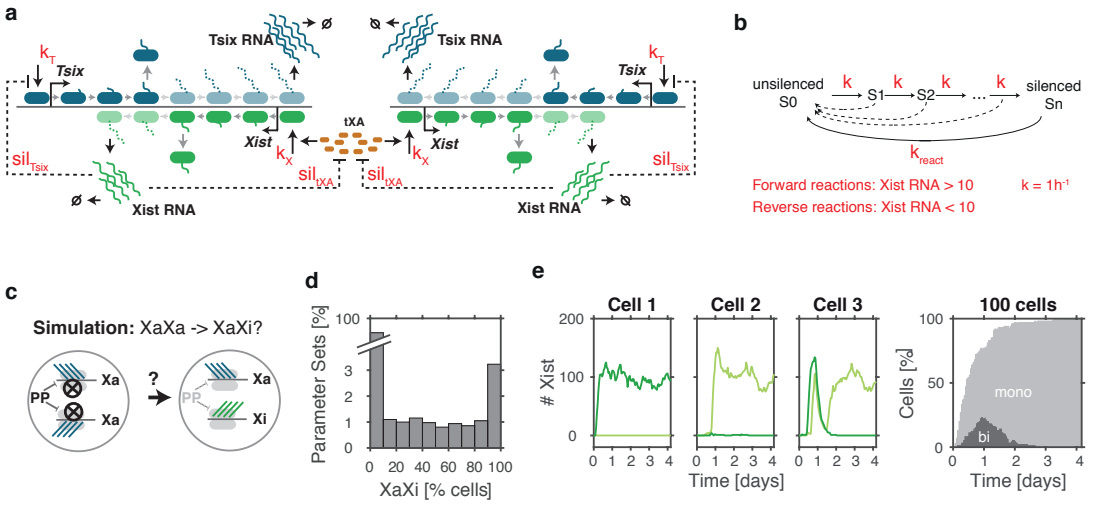

FIGURE 2.17: The Xist/Tsix model can reproduce monoallelic Xist upregulation. a) Schematic depiction of the stochastic Xist/Tsix model, where the two X chromosomes are now coupled together by the tXA that increases the Xist initiation rate k_X. Silencing delays were added that determine how long it takes the Xist RNA to silence tXA (sil_{tXA}) and Tsix (sil_{Tsix}). b) Implementation of the silencing delay: We assumed that once Xist RNA was present with more than 10 molecules, the *Tsix* and *tXA* alleles on that chromosome would stochastically transit through a number of intermediate states until they would reach the final state, which is when their silencing would occur. If the level of Xist RNA molecules dropped below the threshold of 10 again, silencing of Tsix and tXA could be reversed with rates $k_{react-T}$ and $k_{react-tXA}$, respectively. c) Scheme of the simulation. A cell initiates from the undifferentiated XaXa state, where Xist (green) is repressed by pluripotency factors (PP) and Tsix (blue) is transcribed. d) Distribution of the percentage of cells that end up in the XaXi state across all parameter sets. e) Example simulation of a monoallelic parameter set. Left: Three individual cells are shown (Xist1: dark green, Xist2: light green). Right: Percentage of Xist monoallelic (light grey) and biallelic (dark grey) cells in a population of 100 simulated cells is shown.

TABLE 2.11: Silencing delays and reactivation rates for the XaXa → XaXi simulation

Description	Parameter	Parameter Values
Silencing delay of tXA [h]	sil_{tXA}	1 - 48 (log distributed)
Silencing delay of Tsix [h]	sil_{Tsix}	1 - 48 (log distributed)
Reactivation rate of tXA [h^{-1}]	$k_{react-tXA}$	0.1 - 100 (log distributed)
Reactivation rate of Tsix [h^{-1}]	$k_{react-T}$	0.1 - 100 (log distributed)

and let it change its value abruptly to $k_X^{eff} = 1 \cdot (tXA_1 + tXA_2) \cdot k_X$ at the onset of differentiation. To implement the silencing delays for Tsix and tXA, we assumed that Xist RNA, if present above a threshold level of 10 molecules, would induce the stochastic transition of *Tsix* and *tXA* promoters through a number of intermediate states until they reached the final silenced state, at which point they would be silenced in a stepwise manner (Fig 2.17b). If the Xist RNA level falls below the threshold required for silencing, Tsix and tXA can get reactivated with rates $k_{react-tXA}$ and $k_{react-Tsix}$ (Fig 2.17b).

We combined each parameter set that was able to maintain the XaXi state in the previous simulation with 500 randomly sampled values for the Tsix and tXA silencing delays, and for the reactivation rates, with half-lives of the OFF state ranging seconds to several hours (Table 2.11). 100 cells were simulated with each parameter set for a timespan of 4 days, with both of their X chromosomes initiating from the Xa state (Fig 2.17c). Fig 2.17d shows the fraction of monoallelic cells across all tested parameter sets. Approximately 1% of all tested sets robustly upregulated Xist monoallelically in >99% of cells, and were classified as monoallelic (example

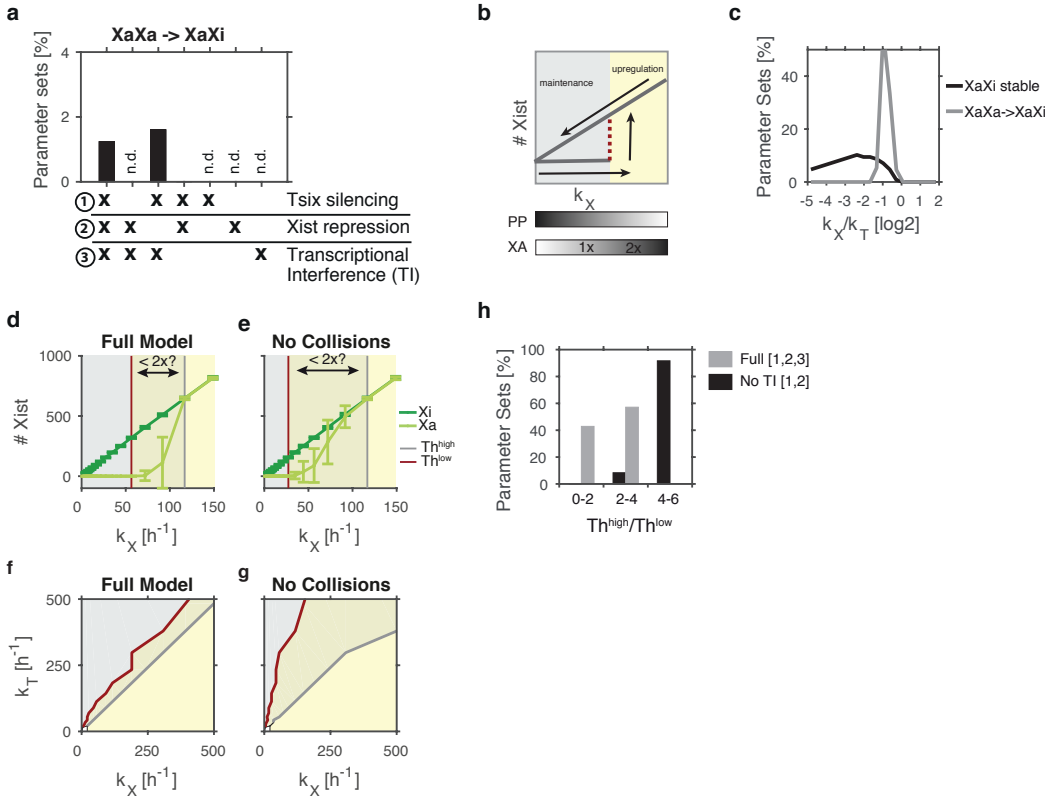

FIGURE 2.18: Reduced model versions. a) The percentage of parameter sets that initiate monoallelic Xist upregulation in >99% of cells is shown for the full model and the reduced models (only those reduced models that could maintain the XaXi state were simulated, n.d.: not determined). b) Schematic depiction of Xa and Xi stability in response to changes in k_X. Bars below indicate how k_X is modulated by pluripotency factors (PP, repression) and tXA (activation). c) Distribution of the k_X-to-k_T ratio in all parameter sets that stably maintain the XaXi state (black) and that allow monoallelic Xist upregulation (grey). d-e) For one example parameter set (k_T=113 h^{-1}, $k_{rev-rep}$=1 h^{-1}) Xist expression (mean and standard deviation) from the X that initiated as Xa (light green) and the X that initiated as Xi (dark green) are shown as a function of k_X for the full model (d) and the model without polymerase collisions (e). The lines indicate the k_X threshold value above which >1% (red, Th^{low}) and >99% (grey, Th^{high}) of cells upregulate Xist from the Xa. f-g) Regimes in which the XaXi state is maintained (grey), or Xist is upregulated from the Xa (yellow) in >99% cells in the k_X-k_T space for the full model (f) and the model without polymerase collisions (g). Red and grey lines indicate the activation thresholds as defined in (d-e). h) Distribution of the Th^{high}-to-Th^{low} ratio (red and grey lines in d-e) in all parameter sets of the full model (grey) and the model without polymerase collisions (black). tXA silencing from the Xi reduces k_X two-fold. Therefore a threshold ratio <2 is required to allow reliable Xist upregulation with a double tXA dose (yellow in d-e), and stable XaXi maintenance with a single tXA dose (grey in d-e). Such a low Th^{high}-to-Th^{low} ratio only exists in the full model.

simulations in Fig 2.17e). Thus the Xist/Tsix model is able to reproduce random choice of an Xi and stable maintenance of the XaXi state.

We again sought to understand which of the repressive mechanisms between Xist and Tsix were strictly required for monoallelic Xist upregulation. To this end, we simulated XCI onset with all reduced model versions that were capable of maintaining the XaXi state in the previous simulation. While Tsix-mediated repression of the *Xist* promoter was still dispensible, we found that RNAP collisions were strictly required for monoallelic upregulation (Fig 2.18a, more on this in section 2.6.3).

We then used the simplified model without Tsix-mediated silencing of the *Xist* promoter to understand which parameter sets allowed monoallelic Xist upregulation. In the presence of a single

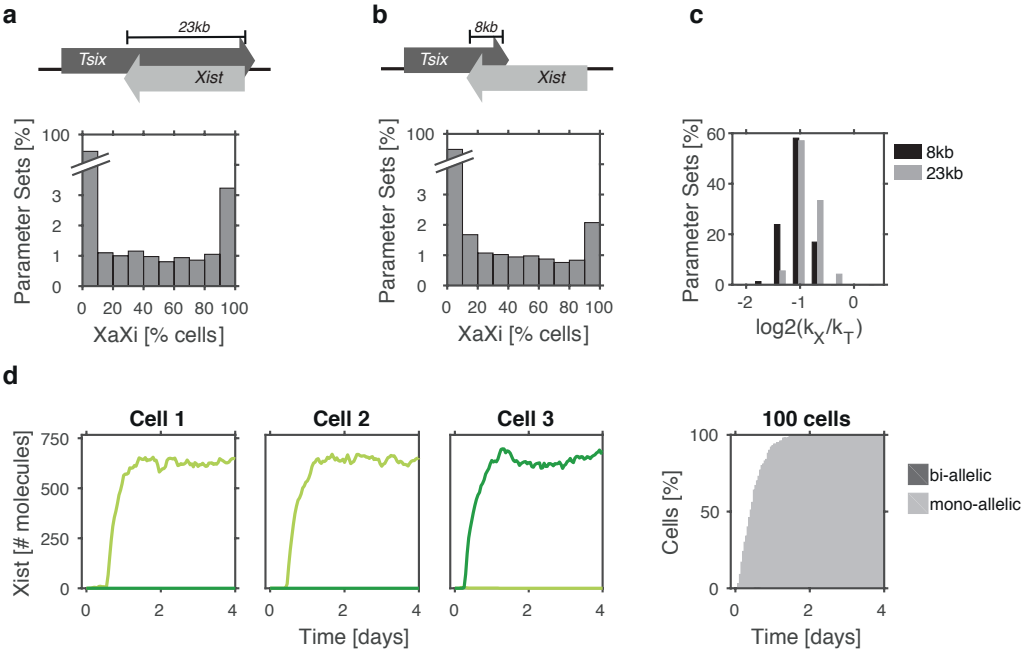

FIGURE 2.19: Reduced overlap between *Xist* and *Tsix* as in the human locus also allows monoallelic upregulation of Xist. a-b) Top: Schematic representation of the *Xist/Tsix* locus in mouse (a) and human (b). Bottom: Distribution of the percentage of monoallelic cells across all simulated parameter sets in the mouse (a) and human (b) locus architecture. c) Distribution of the k_X-to-k_T ratio in the monoallelic parameter sets with the human (black) and mouse (grey) locus architecture. d) Simulation of monoallelic Xist upregulation in one example parameter set with the human locus architecture. Left: Xist trajectories in three individual cells are shown (Xist1: light green, Xist2: dark green). Right: Percentage of Xist monoallelic (light grey) and biallelic (dark grey) cells in a population of 100 simulated cells.

tXA dose the cell must reside below the activation threshold to allow stable XaXi maintenance (Fig 2.18b, grey), but it must cross the activation threshold in the presence of a double tXA dose to trigger Xist upregulation in the XaXa state (Fig 2.18b, yellow). This can only be achieved by parameter sets that lie close enough to the threshold in presence of 1x tXA. Indeed, the k_X-to-k_T ratio of the monoallelic parameter sets was restricted to values between 0.4 and 0.8, such that with a doubling of k_X they would pass the activation threshold that lies slightly below $k_X/k_T = 1$ (Fig 2.18c, grey). In addition, the Xist response to changes in tXA concentration must be switch-like enough, to mediate the required all-or-nothing response within a two-fold change in tXA concentration (Fig 2.18d): In presence of 1x tXA no Xist should be expressed from the Xa, while in presence of 2x tXA only the high Xist expression state should be stable to trigger Xist upregulation in XaXa cells. The vertical lines in Fig 2.18d indicate the threshold value above which >1% (red, Th^{low}) or >99% (grey, Th^{high}) of cells upregulate Xist from the Xa. To allow reliable monoallelic Xist upregulation in pre-XCI cells and stable XaXi maintenance in post-XCI cells, a two-fold reduction in tXA concentration must switch the regime from above the grey threshold to below the red threshold. Interestingly, when comparing the alternative reduced models, we found that such a sharp threshold can only be generated in the presence of polymerase collisions (compare Fig 2.18d,e). In the model without transcriptional interference, all parameter sets required a more than two-fold increase in k_X between the low and the high threshold (Fig 2.18h). This explains why the reduced model without polymerase collisions could not produce monoallelic Xist upregulation.

In the previous section we had found that a reduced overlap between Xist and Tsix was also

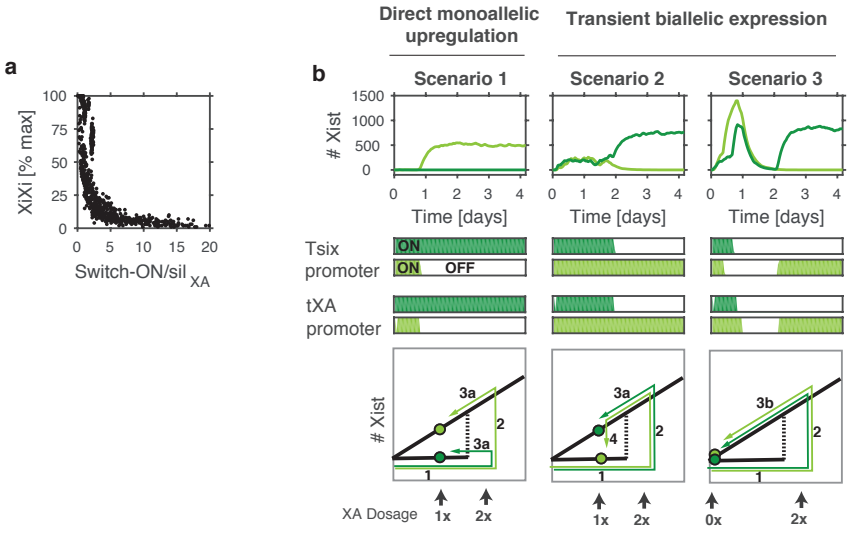

FIGURE 2.20: Different routes to the monoallelic Xist expression state. a) Maximal fraction of biallelic cells observed during the simulation as a function of the ratio of Xist switch-on to tXA silencing delay. The Xist switch-on time was defined as the first timepoint when >10 Xist RNA molecules have been produced. b) Example simulations for three alternative scenarios by which the monoallelic expression state can be reached. Xist expression (top), *Tsix* promoter states (middle), and *tXA* promoter states (bottom) are shown. X1: dark green, X2: light green. Bottom: For each of the three scenarios, a scheme is shown of how the cell's k_X value and Xist expression level changes over the course of diffentiation due to pluripotency factor downregulation (1), Xist upregulation (2), monoallelic (3a) or biallelic (3b) tXA silencing, or Xist switch-off prior to Tsix silencing in the presence of 1xtXA (4).

capable of maintaining the XaXi state. To understand whether such truncated overlap could in principle also generate monoallelic Xist upregulation, we simulated a reduced overlap of 8 kb, similar to the locus architecture reported in human embryonic carcinoma cells. Even with the reduced overlap between XIST and TSIX we found parameter sets that could robustly generate monoallelic XIST upregulation (Fig 2.19b). One example simulation is shown in Fig 2.19d. The monoallelic sets with the 8 kb overlap tended to have lower k_X-to-k_T ratios than the monoallelic parameter sets with the full 23 kb overlap (see Fig 2.19c). For the following analyses we return to the 23 kb overlap model without Tsix-mediated silencing of the *Xist* promoter.

Just as for the general cXR-tXA model we again found parameter sets that could robustly generate monoallelic upregulation but went through a transient period of biallelic Xist expression (scenario 2 and 3, Fig 2.20b). Here too, the extent of transient biallelic expression was determined by the relative timescales of Xist upregulation and tXA silencing (Fig 2.20a). Actually the Xist/Tsix implementation of the positive feedback even predicts that a very low level of biallelic expression (<5%) can be achieved only if switch-on of Xist is at least ∼5 times slower than tXA silencing (Fig 2.20a). Since silencing of X-linked genes takes several hours to days, the switch-on would need to be very slow, which is incompatible with the timing of random XCI in mouse embryos where the transition to the monoallelic expression state occurs within 24 h between E4.5 and E5.5 [142, 155, 156]. The prediction that it is unfeasible to entirely prevent biallelic Xist expression is consistent with our above-described finding that in mouse, too, at least 15% of cells transit through a biallelic expression state.

Fig 2.20b summarizes the alternative routes to monoallelic Xist upregulation. In scenario 1, the symmetry break occurs during Xist upregulation and the monoallelic state is quickly locked

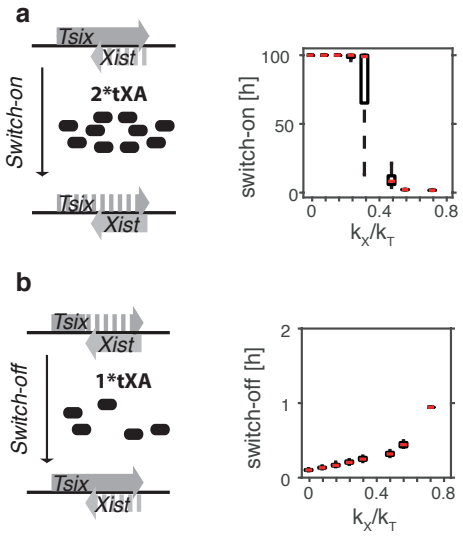

FIGURE 2.21: Two characteristic timescales determining the probability for stable monoallelic Xist upregulation. a) Box plots of the switch-on time (first timepoint, where >10 Xist molecules have been produced) for different k_X-to-k_T ratios. b) Box plots of the switch-off time (first timepoint, where polymerase occupancy on *Tsix* > *Xist*, starting from Xi state in the presence of a single tXA dose) for different k_X-to-k_T ratios. In each box, the center line indicates the median, the box limits indicate the lower and upper quartiles, the whiskers extend to the most extreme data points that were not considered outliers.

in by Tsix and tXA silencing. In scenarios 2 and 3, Xist is biallelically upregulated and the symmetry break must occur during silencing or reactivation. In scenario 2, monoallelic silencing of tXA and Tsix on the future Xi destabilizes Xist expression on the other X where Xist is still repressed by Tsix, and thereby breaks the symmetry. In scenario 3, both X chromosomes are silenced which eventually leads to Xist downregulation. Upon reactivation of tXA and Tsix, the cell can then undertake a second attempt for monoallelic Xist upregulation.

In the following section, I attempt to dissect in which scenarios Xist upregulation will fail. We will start by defining two key timescales that describe the kinetics with which the Xist gene switches between the repressed and transcribed state (Fig 2.21):
(i) The switch-on time measures the time necessary for Xist to switch on transcription and produce >10 molecules in the presence of a double tXA dose. For each parameter set the average switch-on time across all simulated cells was calculated. In cells where switch-on did not happen, it was set to the total simulation time (100h). The switch-on time correlates negatively with the k_X-to-k_T ratio (Fig 2.21a). Thus, the stronger the *Xist* promoter is relative to the *Tsix* promoter the faster the switch-on occurs.
(ii) The switch-off time is defined as the time it takes for the *Xist* locus to switch back from a transcribed to a repressed state, after tXA but before Tsix silencing (Fig 2.21b). Using the maintenance model (section 2.6.1) with a constant tXA dose of one, 100 alleles were simulated for 500 h per parameter set, starting from an Xist transcribing state with the *Tsix* promoter in the ON state. The switch-off time was defined as the first timepoint at which RNAP occupancy on the Xist gene was lower than on the Tsix gene within the overlapping region. Again, if this did not occur before the end of the simulation, the switch-off time was set to the total simulation time. The mean switch-off time correlates positively with the k_X-to-k_T ratio but Xist switch-off happens at much faster timescales than its switch-on (Fig 2.21b). This indicates that Xist expression remains unstable until Tsix has been silenced in *cis*.

We identified six alternative scenarios in which Xist upregulation fails, which are summarized in Figure 2.22 together with an example simulation for each scenario. In brief, monoallelic Xist upregulation requires a precisely controlled k_X-to-k_T ratio, that must be small enough to stabilize

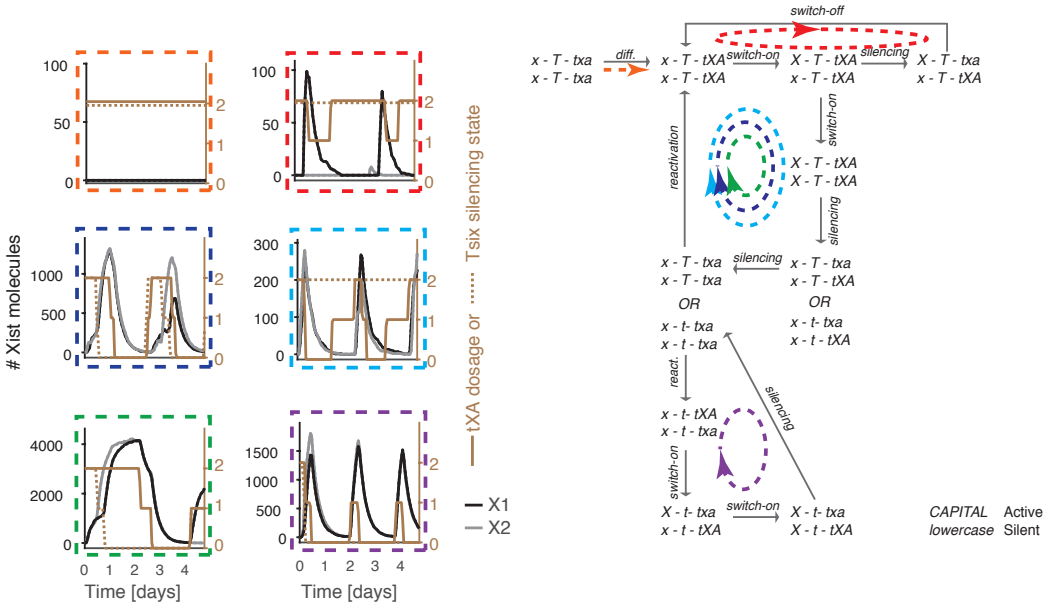

FIGURE 2.22: Scenarios in which monoallelic Xist upregulation fails. Right: Schematic representation of the distinct scenarios that prevent monoallelic Xist upregulation (colored arrows). X: Xist, T: Tsix, tXA: tXA, CAPITAL: active, lowercase: silent. Left: One example simulation is shown for each scenario.

the $Xist^{low}$ state on the Xa in presence of a single tXA dose (see section 2.6.1) but also high enough to pass the activation threshold and ensure fast enough Xist upregulation in the presence of a double tXA dose. This fails in the orange scenario in Fig 2.22, where Xist switch-on is too slow to occur within the relevant timescale. The silencing kinetics determine whether and when a symmetry break between the two X chromosomes occurs. If tXA silencing is faster than Xist switch-on, Xist remains monoallelically upregulated. This monoallelic expression state can still be destabilized if Tsix silencing is too slow (red scenario Fig 2.22). If Xist switch-on is faster than tXA silencing, Xist is upregulated from both alleles and the symmetry between the two X chromosomes must be broken during the silencing or reactivation process. This is difficult to achieve if Tsix and tXA silencing or reactivation occur on very different timescales. In the dark blue and purple scenarios, tXA reactivation occurs too slow or too fast, respectively, compared to Tsix silencing. In the green scenario, tXA silencing is so slow that cells cannot undergo a second attempt to upregulate Xist monoallelically within the relevant timescales. In the light blue scenario, Tsix fails to be silenced. These last scenarios trap cells in "oscillations" of biallelic Xist upregulation and biallelic Xist RNA degradation. In summary, monoallelic Xist upregulation requires a precisely controlled k_X-to-k_T ratio and a symmetry breaking event between the two Xist alleles. The silencing kinetics of tXA and Tsix determine whether and when the break of symmetry occurs. If Tsix and tXA are silenced on very different timescales a symmetry break is prevented.

The precise parameter rules that we used to predict whether a set will generate monoallelic Xist upregulation can be found in Appendix A. Using these rules, we classified the simulated parameter sets as monoallelic or non-monoallelic. Figure 2.23 summarizes the distribution of monoallelic expression among all simulated sets (left panel) and among the sets that were predicted to be monoallelic (right panel). We correctly identified 99% of all non-monoallelic parameter sets as non-monoallelic, but we falsly classified 19% of all monoallelic parameter sets as non-monoallelic.

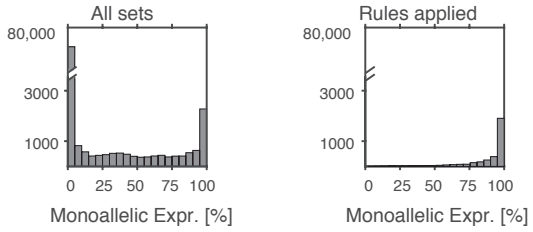

FIGURE 2.23: Prediction efficiency of monoallelic Xist upregulation by applying the parameter rules depicted in Fig 9.2 in the Appendix. Fraction of monoallelic cells in all simulated parameter sets (left) and in all sets that were classified as monoallelic (right).

Among the sets that were classified as monoallelic on average 88% of cells indeed upregulate Xist monoallelically, and in 66% of monoallelically classified sets >90% of cells end up with monoallelic expression (Fig 2.23).

In summary, the Xist/Tsix model can thus recapitulate monoallelic Xist upregulation. Similar to the universal cXR-tXA model, cells can reach the monoallelic expression state via direct monoallelic or transient biallelic Xist upregulation. Importantly, though, mutual repression between Xist and Tsix can only produce monoallelic upregulation of Xist, if we assume the presence of transcriptional interference between the two genes. In the next section, we set out to validate this essential assumption of the Xist/Tsix model experimentally.

2.6.3 Xist represses Tsix by transcriptional interference

The Xist/Tsix model can only produce monoallelic Xist upregulation if transcription interference is assumed to occur between the two genes. Here, we set out to find experimental evidence for this central model assumption. To this end, we made use of a model system that allows us to artificially induce Xist transcription and assess its effect on Tsix transcription independent of Xist RNA-mediated silencing of Tsix. In the absence of Xist RNA-mediated silencing, Tsix transcription within the overlapping region but not upstream of the overlap should be reduced upon Xist induction, if the two genes indeed repress each other by transcriptional interference.

We used two male mESC lines, that both carry a doxycycline-inducible *Xist* promoter (TXY and TXYΔA, Fig 2.24a,b). The TXYΔA line additionally carries a deletion of the repeat A region within Xist, which is essential for Xist RNA-mediated silencing (Fig 2.24b, [12]). Thus, in this line, any repressive effect exerted by Xist on Tsix is not caused by Xist RNA-mediated silencing of Tsix. Due to the switch-like Xist response, the model predicts that at an individual allele either Xist or Tsix will be transcribed at any given time in a mutually exclusive manner. Only very rarely will they be transcribed simultaneously, in a transient state when collisions occur somewhere in the middle of the overlapping region. To obtain single allele measurements of Xist and Tsix transcription, we performed RNA FISH with intronic strand-specific oligo probes. We used a single probe set to detect Xist transcription, and two different probe sets to detect Tsix transcription, one upstream of the overlap with *Xist* (5') and one within the overlap (3') (Fig 2.24c). The doxycycline-inducible *Xist* promoter allowed us to force Xist upregulation. We could then assess whether Xist transcription exerted a repressive effect on Tsix transcription.

In the wild-type TXY line, Xist and Tsix transcription were indeed mutually exclusive on almost all alleles after one day of doxycycline treatment (Fig 2.24d). Strikingly, also in the TXYΔA

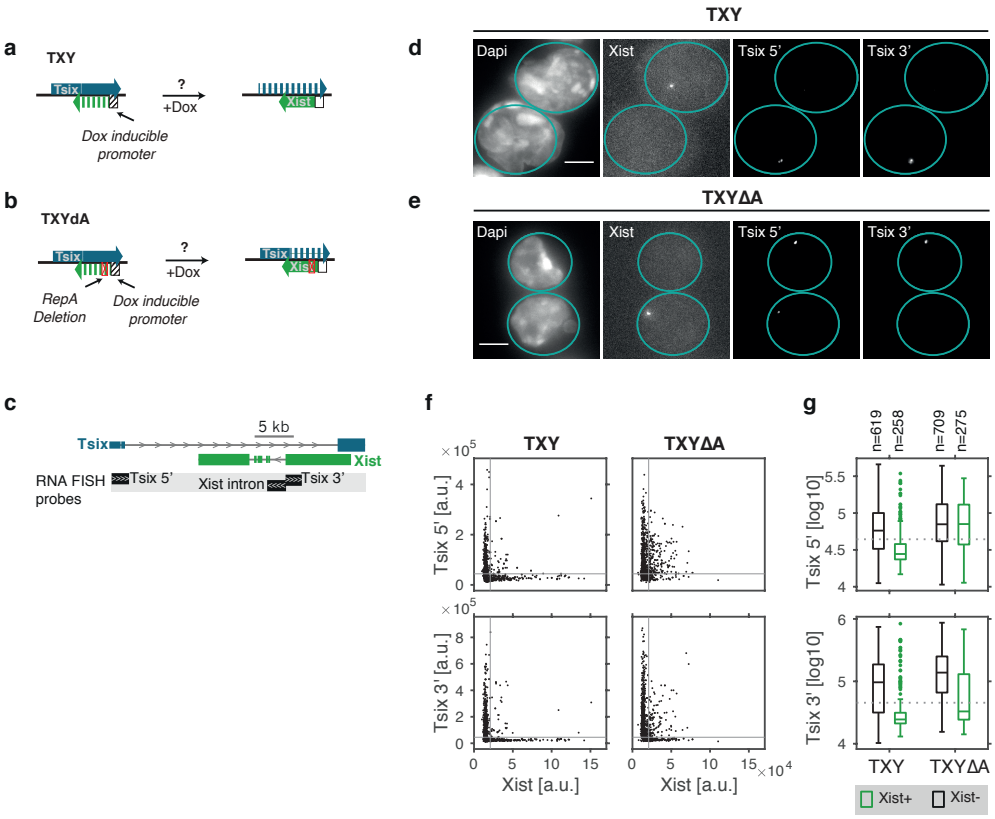

FIGURE 2.24: Transcriptional interference at the *Xist/Tsix* locus. a-b) Scheme of the TXY and TXYΔA lines that both carry a doxycycline-inducible *Xist* promoter (black), allowing us to analyze the effect of enforced Xist transcription on Tsix transcription. The TXYΔA line also carries a repeat A deletion (red) which abolished Xist RNA-mediated silencing (b). c) Position of the strand-specific RNA FISH probes used in (d-g). d-g) TXY and TXYΔA ESCs were treated with doxycycline for 24h and nascent transcription of Xist and Tsix (3' and 5') was analyzed with RNA FISH. d,e) Example images for the TXY and TXYΔA lines. Blue outlines indicate nuclei, white scale bar, 5μm. f) Quantification of 877 (TXY) and 984 (TXYΔA) cells that were pooled from three replicates. Each dot represents the measured signal intensities of a single allele. Grey lines indicate the detection threshold that was estimated for negative control background regions. g) Box plots of the Tsix 5' (top) and 3' (bottom) signal intensities at Xist+ (green) and Xist- (black) alleles in the TXY (left) and the TXYΔA (right) lines. In each box the center line indicates the median, the box limits indicate the lower and upper quartiles, the whiskers extend to the most extreme data points that were not considered outliers and the points indicate the outliers. Grey dotted lines indicate detection threshold.

line, which allows us to observe transcriptional interference independent of Xist RNA-mediated silencing, Xist and Tsix transcription within the overlap (Tsix 3') but not upstream of the overlap (Tsix 5') was mutually exclusive on almost all alleles (Fig 2.24e). We next quantified the Tsix 5' and 3' signal intensities on the Xist transcribing (Xist+) and non-transcribing (Xist-) alleles as a measure for the relative strength of transcription. In the TXY line, Tsix 5' and 3' signals were both strongly reduced at Xist+ alleles compared to Xist- alleles, likely an additive effect of transcriptional interference and Xist RNA-mediated silencing (Fig 2.24f,g, left). Strikingly, in the TXYΔA line, Tsix 3' signals were also reduced at Xist+ alleles (albeit to a lesser extent than in the silencing competent wild-type TXY), while Tsix 5' signals were barely affected (Fig 2.24f,g, right). The individual replicates of this experiment are shown in Appendix B. Thus, Tsix transcription specifically within the overlap is clearly perturbed in the absence of Xist RNA-mediated silencing, indicating the presence of transcriptional interference and thereby validating a central assumption of the Xist/Tsix model. However, Xist RNA-mediated silencing also contributes to Tsix repression, as Tsix signals are even lower in the TXY line (Fig 2.24g).

2.6.4 The antisense model correctly reproduces phenotypes of Xist and Tsix mutants

As a further test of the antisense model, we simulated Xist and Tsix mutants whose phenotypes had already been assessed experimentally [20, 24, 157]. We simulated four different genotypes: Wild type, Tsix$^{+/-}$, Tsix$^{-/-}$ and Xist$^{+/-}$ (Fig 2.25a-d) with the reduced antisense model where Xist and Tsix mutually repress each other by transcriptional interference and Xist RNA-mediated silencing (see section 2.6.2). We selected a subset of parameters from the previous simulation of monoallelic Xist upregulation that were in agreement with experimental observations using the following constraints:

- Transient biallelic Xist expression: The maximal percentage of cells in a biallelic expression state had to be <20% at any given time, as this is approximately the extent of biallelic expression that we observed in vivo and in vitro (see section 2.4).
- Xist expression level: The mean Xist expression level at the end of the simulation must lie between 200 and 600 RNA molecules [92, 134].
- Timing of Xist upregulation: All cells must have upregulated Xist 48h after induction of differentiation, as the fraction of Xist expressing cells does not increase further after two days, neither in vivo nor in vitro (see section 2.4).

Since we only found 34 sets with unique k_X, k_T, sil_{tXA} and sil_{Tsix} values that fulfilled these criteria, we performed an additional simulation of monoallelic Xist upregulation in the wild type with 500,000 sets to identify more parameter sets that were in accordance with these constraints (see methods section 9.3.1.3 for details). We then randomly selected 100 sets that fulfilled the above criteria to simulate the mutants. To simulate a deletion, we set the initiation rate of that respective allele to zero:
(i) Tsix$^{+/-}$: $k_{T_{X2}} = 0$
(ii) Tsix$^{-/-}$: $k_{T_{X1}} = k_{T_{X2}} = 0$
(iii) Xist$^{+/-}$: $k_{X_{X2}} = 0$

For each parameter set we simulated 100 cells for each of the four genotypes with the stochastic Gillespie algorithm. In wild-type cells, our simulations predict that each of the two X chromosomes is inactivated with equal probability, so that half of the cells end up with Xist expression from one, the other half from the other X (Fig 2.25a, bottom). In agreement with experimental data, our model predicts that heterozygous Tsix or Xist mutants exhibit non-random XCI. In heterozygous Tsix mutants, the mutant X chromosome is always inactivated [20] (Fig 2.25b, bottom), while in heterozygous Xist mutants it is the wild-type X that is inactivated in all cells (Fig 2.25d, bottom). Experimental data also shows that the kinetics of Xist upregulation are affected in these two mutant lines: XCI is accelerated in heterozygous Tsix mutants and slowed down in heterozygous Xist mutants compared to the wild type [107]. To test whether the model recapitulates these observations, we calculated the half-time of Xist upregulation ($T_{1/2}$), where 50% of the cells had switched on Xist and compared these values between wild type and mutants. This is visualized in an example simulation of the same parameter set in the wild type and two heterozygous mutants (Fig 2.25e). Fig 2.25f shows the distribution of the change in half-time in

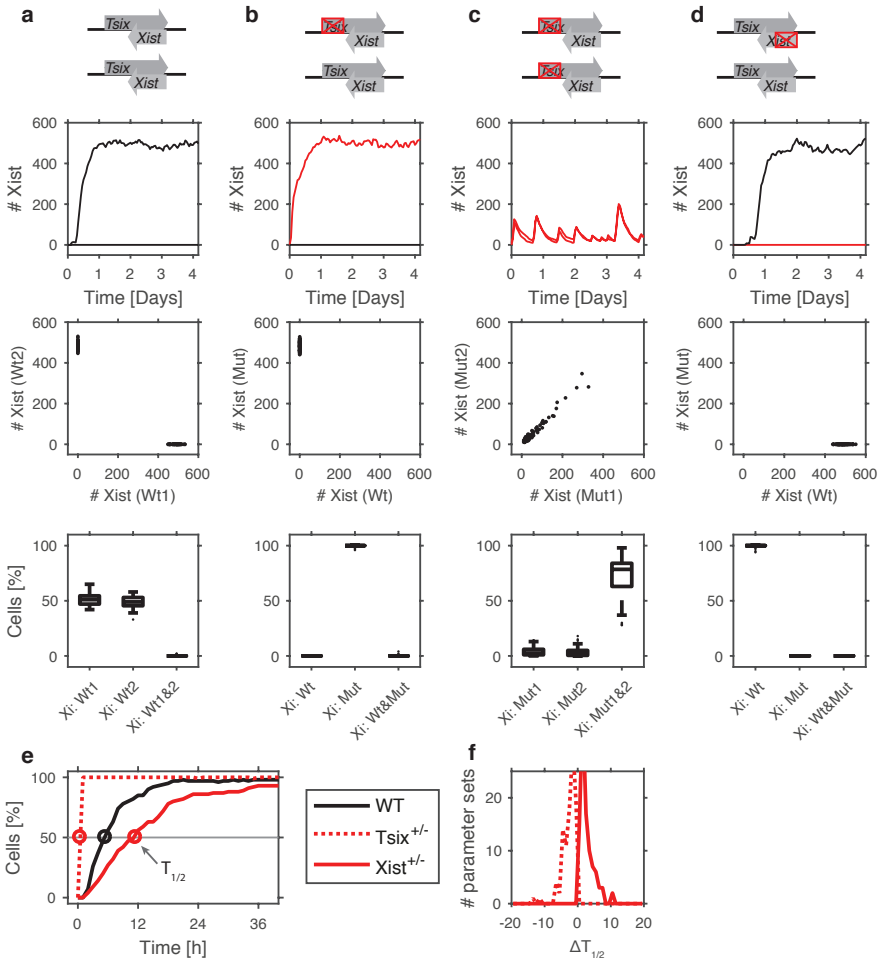

FIGURE 2.25: The model correctly reproduces the phenotypes of *Xist/Tsix* mutants. a-d) Simulation of the *Xist/Tsix* mutant cell lines (top). Representative simulation with a single parameter set of an individual cell showing Xist expression from wild type (black) and mutant (red) alleles (upper middle), and in a population of 100 cells (lower middle). Boxplots depict the percentage of cells that express Xist monoallelically from each of the two alleles or biallelically across all simulated parameter sets (bottom). e-f) Timescales of Xist upregulation in wild type and heterozygous mutants. e) Representative simulation with an individual parameter set depicting the percentage of Xist+ cells over time. The half-time ($T_{1/2}$) at which 50% of cells have upregulated Xist is indicated. f) Distribution of the change in half-time ($\Delta T_{1/2}$) in the mutants compared to the wild type.

the mutants compared to the wild type. Indeed, a *Tsix* deletion reduces the half-time of Xist upregulation, while an *Xist* deletion increases it. As discussed before, the kinetics of Xist upregulation are positively correlated with the k_X-to-k_T ratio, which increases upon *Tsix* deletion. Upon heterozygous *Xist* deletion, only the wild-type allele remains capable of upregulating Xist, which increases the average switch-on time.

The last mutant that we simulated was a homozygous *Tsix* deletion. These mutants have been reported to undergo "chaotic" XCI, with a mixture of cells that inactivate no, one or two X chromosomes. Among the cells with one inactive X, the choice of the Xi is again random [24, 158]. Our simulations predict that a homozygous *Tsix* deletion will result in Xist oscillations (Fig 2.25c, top). Since *Tsix* deletion strongly reduces the Xist switch-on time on both alleles, Xist is quickly upregulated biallelically. Complete tXA silencing results in Xist downregulation, which is immediately followed by another round of biallelic Xist upregulation. This is comparable to the failure of Xist upregulation in the purple scenario that we had identified in Fig 2.25c

(group 1.2.4 A), where cells are trapped in oscillations of biallelic Xist upregulation because tXA is reactivated faster than Tsix. In agreement with the experimental data, the simulations of the Tsix$^{-/-}$ mutant show a higher fraction of biallelically Xist expressing cells (Fig 2.25c, bottom).

We set out to validate the Tsix$^{-/-}$ model prediction more precisely. Specifically, we wanted to find an experimental indication for the predicted oscillations between biallelic Xist expression and complete tXA silencing. To this end, we generated a mESC line with a homozygous *Tsix* deletion using CRISPR/Cas9-mediated genome editing. We used two guides targeting either side of a ∼4 kb fragment containing the *DxPas34* region, that acts as Tsix enhancer (Fig 2.26a). Single clones were expanded and genotyped to identify those that carried the deletion. Sanger sequencing of PCR products was used to verify that the deletion was present on both alleles using annotated SNPs between the B6 and Cast alleles. Correct karyotype of promising clones was also verified. We then assessed Tsix expression in undifferentiated cells at the single allele level using RNA FISH with two different oligonucleotide-based probe sets (Tsix 5' and Tsix 3', Fig 2.26a). The vast majority of oligos in the Tsix 5' set bind within the deleted region. Therefore this probe set allows us to verify the purity of the clones. The Tsix 3' probe binds further downstream within the overlap with Xist and allows us to assess residual Tsix expression, that might still originate from alternative upstream *Tsix* TSSes. While the Tsix 5' signal was basically absent in all three clones, there was still a substantial fraction of cells with a Tsix 3' signal (on average 48% Tsix 3'+ cells compared to 85% in the wild type (WT)), indicating residual Tsix expression (Fig 2.26b). Therefore, the effects we observe are likely milder than predicted by our simulations which assumed complete absence of Tsix transcription. We next assessed Xist expression upon the induction of XCI by differentiation through LIF withdrawal. Compared to the wild type we indeed observed an increased fraction of biallelically Xist expressing cells, which was also sustained for a longer period of time (Fig 2.26c). While three days after induction of differentiation, virtually all cells in the wild type had resolved biallelic to monoallelic expression, the three Tsix$^{-/-}$ mutants consistently maintained between ∼7% and 18% biallelic cells. Although our *DxPas34* deletion did not abolish but only reduce Tsix transcription, we observed an increased fraction of biallelic Xist expressing cells, consistent with our simulations and previous experimental data [24]. These preliminary results must be repeated and confirmed by an alternative quantification method. In addition, the essential experiment to capture oscillations in Xist expression, remains to be done and could be addressed by combining Xist RNA FISH with H3K27me3 immunofluorescence.

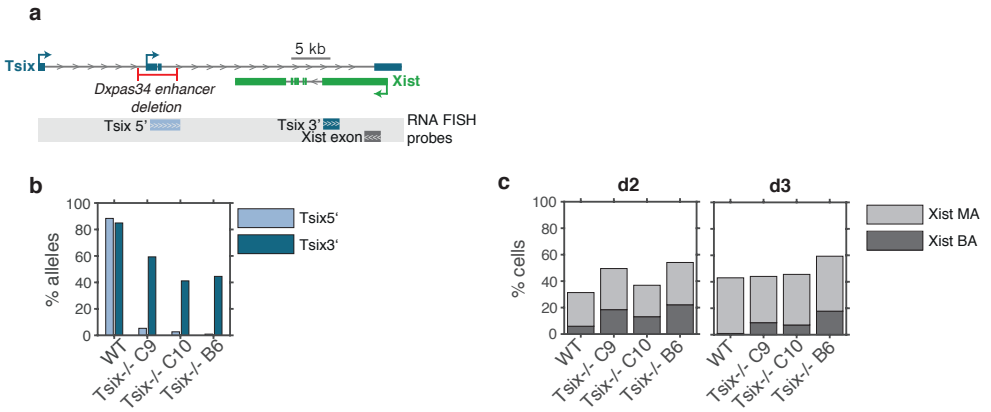

FIGURE 2.26: Homozygous *Tsix* deletion increases the extent of biallelic Xist expression. a) Schematic depiction of the *Xist/Tsix* locus. *DxPas34* enhancer deletion is shown in red and Xist and Tsix (5' and 3') RNA FISH probes are indicated. b) Percentage of Tsix 5' (light blue) and Tsix 3' (dark blue) positive alleles is show in the wild type and the three Tsix-/- clones (n>100). c) Stacked bar graphs show the percentage of Xist monoallelic (light grey) and biallelic (dark grey) cells in wild type and the three Tsix-/- clones at different timepoints after induction of differentiation by LIF withdrawal (n>100 for each timepoint).

2.6.5 Aneuploid and polyploid cells in the Xist/Tsix model

As a final test for the Xist/Tsix antisense model, we verified that it would, just as the cXR-tXA model, correctly recapitulate the patterns of Xist expression in aneuploid and polyploid cells. Again we assumed that each additional X chromosome would produce an additional tXA dose. In X aneuploid cells this results in destabilization of the Xa state in the presence of more than one tXA dose such that indeed all tested parameter sets correctly predict no Xist expression in male and XO cells, and bi- or triallelic expression in X-chromosome tri- or tetrasomies, respectively (Fig 2.27a).

To simulate polyploid cells we again assumed that a two-fold increase in the genome size would result in a two-fold increase in nuclear volume, and therefore in tXA dilution, so that 4n4X and 2n2X cells have the same tXA concentration (Fig 2.27b, left). A tetraploid cell with 2 Xas will thus reside in the bistable regime while with 4 Xas it will reside in the monostable regime (Fig 2.27b, left). As discussed in section 2.3 it is more challenging to distinguish between 2x tXA and 3x tXA dose in tetraploid cells, than between a 1x tXA and 2x tXA in diploid cells, as the switch-like response of Xist must occur within a 50% increase rather than a 100% increase in tXA. Therefore, parameter sets exist in which the switch-like response is not sharp enough and Xa and Xi are both stable in the presence of 3x tXA, making the XiXaXaXa cellular state stable (black arrows Fig 2.27b, left). Indeed, we find that in some sets a considerable fraction of cells ends up with only 1 Xi (light green Fig 2.27b, middle). However, if we only select those sets in which the Xa state is unstable in the presence of a 3x tXA (blue arrows Fig 2.27b, left), they can indeed reliably produce biallelic Xist upregulation in tetraploid cells (Fig 2.27b, right). It should be noted that silencing of the tXA must be sufficiently fast to prevent inactivation of three X chromosomes. The cellular state with XaXiXiXi is also stable, because the Xi remains

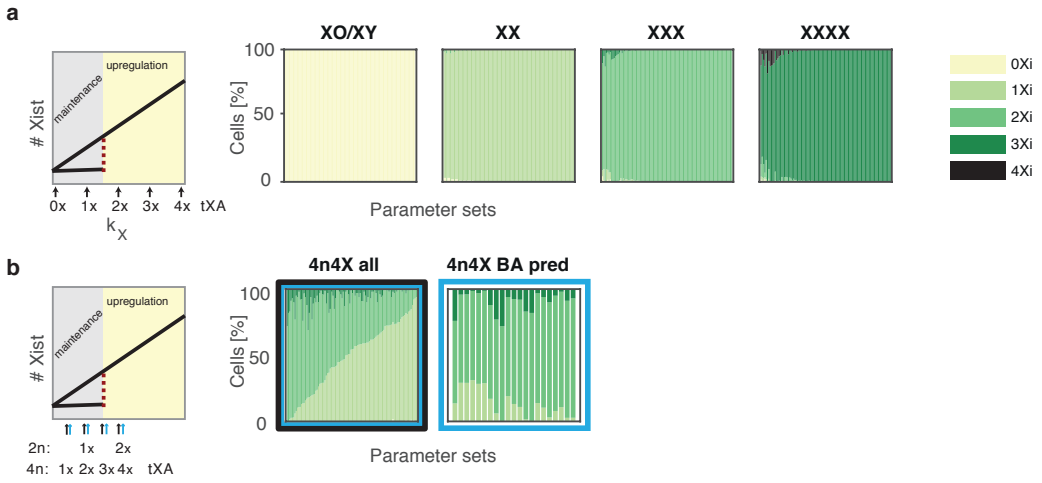

FIGURE 2.27: The Xist/Tsix model can reproduce Xist expression patterns in aneuploid and polyploid cell lines. a) Simulation of diploid cells with 1, 2, 3 or 4 X chromosomes. Left: Scheme of Xa and Xi stability as a function of k_X, which is modulated by the tXA dose. Additional X chromosomes produce additional tXA doses, resulting in destabilization of the Xa (yellow), until only a single tXA dose remains (grey). Right: Stacked bar graphs depict the percentage of cells with 0, 1, 2, 3 or 4 Xis in the different simulated genotypes (after 4d of simulation). A chromosome was classified as Xist+ if >10 Xist RNA molecules were present. b) Simulation of tetraploid cells (4n4X) assuming that a doubling in nuclear volume in tetraploid cells will result in a two-fold tXA dilution. Right: Scheme of Xa and Xi stability as a function of k_X. Arrows below indicate how k_X is modulated by the tXA dose in diploid (2n) versus tetraploid (4n) cells. Right: Stacked bar graphs depict the percentage of cells with 0, 1, 2, 3 and 4 Xis in all parameter sets (black+blue, left) and in those sets (blue, right) where a 3x tXA scheme resides in the upregulation regime (yellow).

stable even in the presence of only a single tXA dose.

2.6.6 Summary of the Xist/Tsix model

We have proposed a molecular candidate for the predicted cXR factor - Xist's antisense transcript Tsix. Tsix represses Xist in *cis* and is itself subject to Xist RNA-mediated silencing. It could therefore implement the required double negative feedback. Using a stochastic model of Xist and Tsix transcription, we have demonstrated that mutual repression between Xist and Tsix can stabilize the XaXi state, and generate monoallelic Xist upregulation, given that Xist and Tsix also repress each other by transcriptional interference, an assumption that we validated experimentally. We also showed that stable monoallelic Xist upregulation is possible with a reduced overlap between Xist and Tsix, as reported in human embryonic carcinoma cells.

Thus, we have shown that Tsix is a promising cXR candidate. Mutual repression between Xist and Tsix could be the mechanism that implements the positive feedback necessary for stable XaXi maintenance, and can in addition tightly control the Xist switch-on time to allow for monoallelic Xist upregulation. The Xist/Tsix model can also recapitulate XCI patterns in aneuploid and polyploid cells and is in concordance with the Xist expression patterns in a number of Xist and Tsix mutant cell lines.

3 Discussion

The establishment of alternative fates for two identical X chromosomes is a fascinating process that scientists have been studying for almost 60 years now. In this work, we set out to identify the simplest model topology that can explain random XCI onset. Specifically, it must allow cells to (1) assess their number of X chromosomes and initiate XCI only if they posses two or more Xs (per diploid set of autosomes), (2) make a mutually exclusive choice for an inactive X, and (3) stably maintain this choice. We have identified the simplest network that can ensure female-specific and monoallelic Xist expression by systematically screening all model topologies with up to two X-linked Xist regulators. The identified network consists of a *cis*-acting Xist repressor and a *trans*-acting Xist activator that are both susceptible to Xist RNA-mediated silencing.

The network correctly predicts the Xist expression patterns of aneuploid and polyploid cells and further makes the novel prediction that cells are able to revert biallelic Xist expression because complete tXA silencing destabilizes the $Xist^{high}$ state. We indeed find the first direct experimental evidence for the ability of cells to revert biallelic Xist expression.

We propose that the predicted cXR could be Xist's antisense transcript Tsix and demonstrate that the antisense locus can mediate the required bistable switch using a detailed mechanistic model of Xist and Tsix transcription. We validate the model's assumption that the two genes repress each other by transcriptional interference using a system that allows us to assess the effect of Xist transcription on Tsix transcription independent of Xist RNA-mediated silencing. The antisense model correctly predicts the phenotypes of several published Xist and Tsix mutant cell lines.

Our work postulates the first ever mechanistic model that describes the complete XCI onset. It correctly reproduces known mutant phenotypes and makes novel predictions that we have verified experimentally. Our mathematical modeling approach is complementary to previous studies as it uncovers the core regulatory interactions that are necessary for XCI onset, rather than identifying individual regulators. This allows us to postulate a minimal XCI network which can serve as a basis to identify the required regulators and the molecular mechanism by which they mediate the predicted regulatory loops. These regulatory interactions could also be implemented indirectly. For instance the tXA could activate Xist by repressing an autosomal Xist repressor (reminiscent of Xist activation through Rnf12-mediated degradation of Rex1), or by activating an autosomal Xist activator in an ultrasensitive manner. Similarily, multiple regulators could act together in a redundant or additive manner.

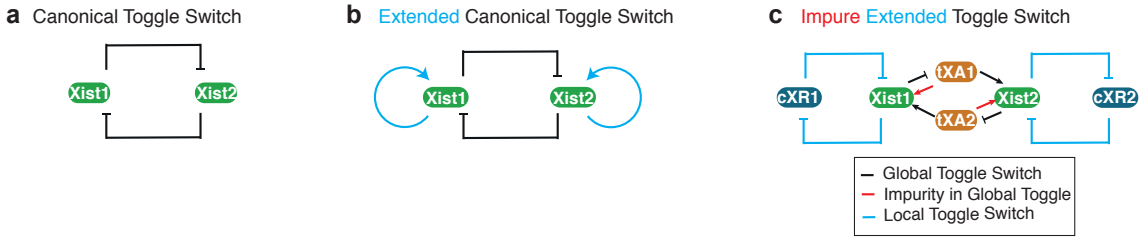

FIGURE 3.1: A modified extended toggle switch can explain XCI onset. a) The canonical toggle switch: Mutual inhibition between two entities as the canonical model for mutual exclusivity in biology. b) The extended canonical toggle switch: Each entity is stabilized by a positive autoregulatory feedback. c) The cXR-tXA circuitry: It differs from the canonical model by the absence of pure reciprocality (red) and by the positive autoregulatory feedback being composed of a double negative feedback (local toggle switch, blue).

3.1 A modified extended toggle switch can explain XCI onset

XCI onset represents a problem of mutual exclusivity, where each female cell must either assume the XaXi or the XiXa state. In biology, problems of mutual exclusivity are often solved by mutual inhibition between two entities in a so called toggle switch (Fig 3.1a) [122]. Such mutual repression can create two alternative stable states, where one entity is active and represses the other and vice versa. In addition, the two entities can be subject to a positive feedback loop, giving rise to the "extended toggle switch" (Fig 3.1b) [124, 159].

How is mutual inhibition between the two Xist alleles, as the fundamental element for mutual exclusivity implemented in our cXR-tXA network? In fact, the two Xist alleles mutually inhibit each other through tXA silencing: The tXA activates Xist in a dose-dependent manner and thereby ensures female specificity. Xist is only upregulated in the presence of two active tXA alleles. Upon Xist upregulation tXA is silenced in *cis* which reduces its dose to that in males and prevents Xist switch-on from the other X, as proposed previously [107, 109]. Thus by silencing the tXA in *cis*, Xist1 inhibits upregulation of Xist2 (black in Fig 3.1c, [122]). The inhibition is however not purely reciprocal: As tXA acts in *trans*, its silencing will affect both Xist alleles. The tXA-mediated global negative feedback thus differs from the canonical toggle switch of mutual exclusivity, because the inhibition between the two Xist alleles is not purely reciprocal (red in Fig 3.1c). This negative feedback of self-inhibition creates the need for positive autoregulation, which is mediated by the second regulator, the cXR. Mutual repression between cXR and Xist implements a local toggle switch (blue in Fig 3.1c, [124, 159]) that is necessary to neutralize the impurity in the reciprocal inhibition and allow for mutual exclusivity. The double negative (=positive) feedback stabilizes Xist expression in the presence of only a single tXA dose, and creates an ultrasensitive response of Xist to changes in the tXA dose. This allows for a sharp switch-like Xist upregulation between a single and a double tXA dose. Thus, our core XCI network consists of only two regulators, a *cis*-acting Xist repressor and a *trans*-acting Xist activator that together ensure stable monoallelic and female-specific Xist upregulation by forming a modified extended toggle switch (Fig 3.1c).

To allow for robust maintenance of alternative Xist expression states on the Xa and the Xi, each Xist allele must exhibit a memory of its previous expression state. Any bistable circuit

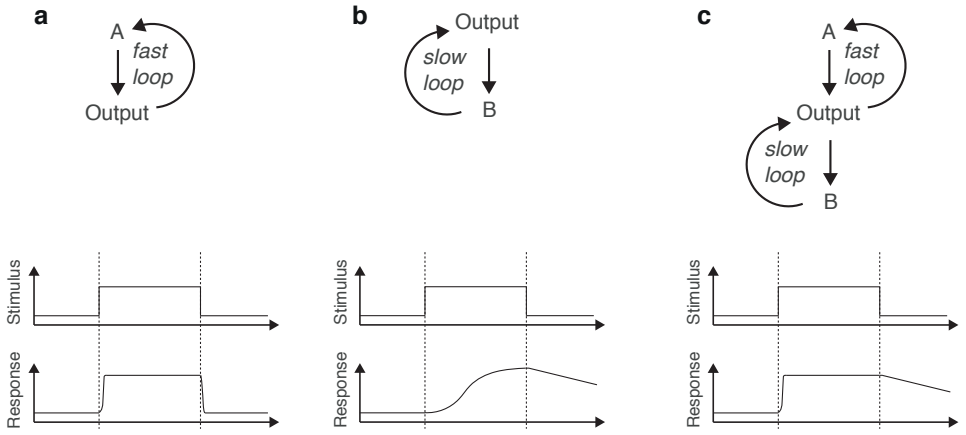

FIGURE 3.2: Coupled feedbacks can ensure fast and reliable decision making. Schematic of feedbacks and feedback response curves over time for a single and two positive feedbacks as modeled in [160]. a) A single fast feedback. b) A single slow feedback. c) One fast and one slow feedback. Figure adapted from [160].

usually exhibits some degree of hysteresis, meaning that the system responds differently to the same stimulus depending on whether it was previously in its ON or OFF state [123]. The two stable states of the system must also be sufficiently robust against stochastic fluctuations. In an artificially designed bistable switch in *S. cerevisiae* Becskei and colleagues found that stochastic fluctuations can cause the toggle switch to flip between states. They had constructed a GFP transcription unit that stimulates its own transcription in the presence of doxycycline. At intermediate doxycycline concentrations individual cells sometimes switched between the ON and the OFF state, likely because stochastic fluctuations within individual cells were sufficient to trigger a flip between states [161]. Many endogenous biological systems that create all-or-nothing outputs contain not only a single but multiple positive feedback loops [160]. One example is the *Xenopus* oocyte maturation, where Cdc2 activates itself via three distinct positive feedback loops and thereby drives entry into mitosis [162]. Two of these feedbacks act through phosphorylation and dephosphorylation on very rapid timescales, while the third feedback between Cdc2 and the MAPK cascade involves translation and therefore acts on a much slower timescale. Another "multiple-loop" example is polarization in budding yeast, which also depends on a rapid and a slow feedback [163]. When the slow positive feedback is blocked, cell polarization happens more rapidly but is less stable. When the fast feedback is blocked, polarization is more stable but slower [163]. This suggests that the slow positive feedback is crucial for stability of the ON state, and the fast positive feedback is critical for the speed of the transition from the OFF to the ON state, a hypothesis that was tested computationally by Brandman and colleagues [160] (Fig 3.2).

One could hypothesize that the reason why we strictly require Xist RNA-mediated silencing for robust XaXi maintenance, is that it increases the system's ability to filter out noise as it acts on a much slower timescale than the rapid polymerase-mediated transcriptional interference. The fast-acting interference on the other hand might be required to speed up Xist upregulation, consistent with our finding that in the absence of collisions the transition region, in which only a fraction of cells upregulated Xist, broadened considerably. Potentially, the Xist/Tsix-mediated positive feedback even acts together with yet another positive feedback to generate even stronger robustness against noise, resulting for instance from stochastic fluctuations in tXA expression,

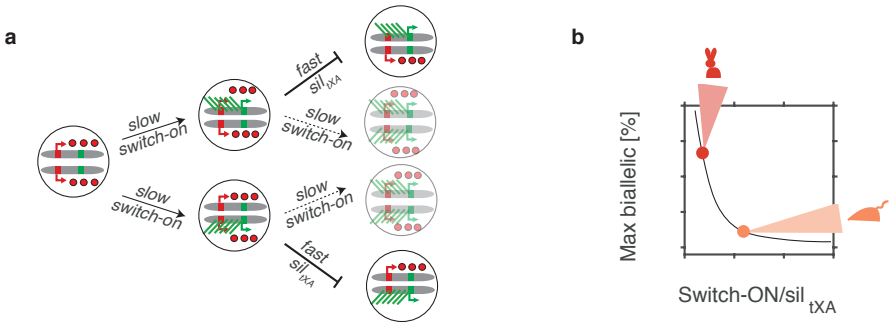

FIGURE 3.3: Timescale separation between Xist upregulation and tXA silencing is required for monoallelic Xist upregulation. a) Xist switch-on must be slow and stochastic to guarantee that it occurs on one allele at a time, tXA silencing must be fast to prevent Xist upregulation from the second allele. b) Different species make different compromises in the speed/reliability trade-off. In mice, slow Xist switch-on guarantees precise monoallelic Xist upregulation (orange), in rabbits faster Xist upregulation comes at the expense of higher transient biallelic Xist expression (red).

which we neglected in our antisense model.

A property of the network that is essential for generating robust monoallelic Xist upregulation is the timescale separation between the initial Xist upregulation and the silencing-mediated negative feedback (Fig 3.3a) [164]. The initial Xist switch-on must be slow to prolong the period between its upregulation from the first and the second allele. This ensures that it will generally occur at one allele at a time, breaking the symmetry between the two Xist alleles. The negative feedback of tXA silencing must be faster to ensure that monoallelic choice is locked in prior to reoccurence of Xist upregulation from the second allele (Fig 3.3a). More precise monoallelic Xist switch-on thus comes at the expense of a slower XCI process. The negative feedback permits some degree of imprecision as it can correct erroneous biallelic Xist upregulation and thereby weaken the requirement for a strict timescale separation. Different mammals seem to make different compromises in this speed/reliability trade-off: While mice seem to go for precise monoallelic Xist upregulation, rabbits accept a higher degree of transient erroneous biallelic Xist expression that likely gets corrected to monoallelic Xist expression by the negative feedback (Fig 3.3b) [99, 141, 165, 166].

These considerations suggest that tXA should be a rather quickly silenced X-linked gene to guarantee a fast negative feedback. Slow Xist switch-on is likely ensured by different factors: On the one hand Xist repression by cXR will reduce the speed of Xist upregulation, as we have demonstrated with the simulation of a heterozygous *Tsix* deletion, which speeds up Xist switch-on in *cis*, in agreement with experimental observations [107]. On the other hand, *trans*-acting factors can contribute to the timescale separation by repressing Xist. For instance, one could speculate that one of the reasons why differentiation is halted in female cells until dosage compensation has been achieved, is that higher levels of pluripotency factors are required to ensure proper timing of Xist switch-on [58, 63, 146, 147, 167]. In male cells, pluripotency factors can be downregulated quickly because other (X-dosage-mediated) mechanisms are in place that ensure a female-specific window of opportunity for Xist upregulation, such as 50% lower tXA expression and potentially faster global DNA methylation [56, 58, 146, 147].

In summary, the global negative feedback of tXA silencing ensures female specificity and mutually exclusive choice, the positive feedback ensures stable XaXi maintenance and creates the threshold required for tXA-mediated female specificity. Other X-linked genes likely mediate global effects that contribute to ensuring correct XCI onset [147].

Interestingly, a similar combination of a local positive and a global negative feedback seems to ensure mutually exclusive expression of a single receptor during olfactory neuron differentiation. Here, each cell must make a stochastic but stable choice for expressing a single one out of 1000 different olfactory receptors. The activation of a receptor gene triggers a global negative feedback that ensures the mutually exclusive choice, and the decision is locked in by a local positive feedback mediated by nucleosome modifications [168–170].

3.2 Role of the cXR - threshold response, timescale separation and XaXi maintenance

Our model screen has shown that cXR is the only X-linked Xist regulator that can maintain the XaXi state, consistent with other biological examples where bistability arises from positive or double negative feedbacks [122, 161, 171]. By blocking the cXR-mediated interaction, we have shown that in our model cXR is required to stabilize both allelic Xist expression states on the Xi and Xa. It allows the cell to memorize alternative Xist expression states, until locked in by epigenetic mechanisms such as DNA methylation.

In addition to stabilizing alternative expression states on Xa and Xi, the local toggle switch between Xist and cXR contains a non-linearity that ensures switch-like or ultrasensitive Xist activation. It allows to convert a twofold difference in tXA levels between males and females into a binary decision, where XCI is initiated in females only. Close to the threshold, a small change in input signal (tXA) results in a large change in the response (Xist).

cXR also plays another important role during XCI onset where the cell has to transit from the XaXa to the XaXi state. During this period, cXR prolongs the switch-on time of Xist and thereby guarantees that Xist switch-on is slow and stochastic, occuring on one allele at a time. In this way, a time window is generated in which the negative feedback can set in and prevent biallelic Xist upregulation. We have experimentally validated that the extent of biallelic expression is controlled by the switch-on time by inducing one of the two Xist alleles with doxycycline. This artifically prolongs the period between switch-on of the first and second Xist allele, and indeed results in a lower fraction of biallelic cells. As a consequence our network can produce different extents of biallelic Xist expression depending on the relative timescales of switch-on to tXA-silencing. This suggests that the XCI regulatory network could be conserved between species although they might exhibit different degrees of transient biallelic Xist expression. Under the assumption that cXR is dampened or only upregulated later, the network can even recapitulate extented biallelic Xist expression, as observed in humans [99, 100].

Our analysis predicts that the cXR must be a rather quickly silenced gene, because the positive feedback must quickly memorize the high state to stabilize Xist expression in presence of 1x tXA. Additionally, its heterozygous deletion should skew Xi choice by reducing the switch-on time in *cis*. It is likely that the cXR factor is a lncRNA locus, because they oftentimes act in *cis* and in addition require the act of transcription or even the transcript itself to exert their function [172]. They are, therefore, susceptible to Xist RNA-mediated silencing. For other obvious *cis*-acting factors, such as DNA elements it is less straightfoward to argue that their repressive effect on Xist would be impaired by Xist RNA-mediated silencing.

Linx is one lncRNA locus within the Xic that represses Xist in *cis* [19]. However it has been shown to exert its function independent of transcription [34]. A more promising candidate for the *cis* repressor is Xist's antisense transcription unit *Tsix*, which overlaps the complete *Xist* gene including its promoter [20, 173]. Tsix transcription establishes a repressive chromatin state at the *Xist* promoter [25–28]. Heterozygous *Tsix* deletions skew the choice of the inactive X towards the mutated allele, as expected from a *cis* repressor [20–23]. We have therefore proposed Tsix as a candidate for the cXR factor and have used a mechanistic model to demonstrate that antisense transcription can indeed generate and maintain monoallelic Xist expression. Transcriptional interference, which we have demonstrated experimentally, generates the precise threshold required for switch-like Xist upregulation. The model also correctly reproduces the phenotypes of Tsix mutant cell lines, including the kinetics of Xist upregulation, which are accelerated in heterozygous Tsix mutants and slowed down in heterozygous Xist mutants [107]. This observation is consistent with the role of the cXR factor in controlling the speed of Xist switch-on and thereby modulating the extent of biallelic Xist expression: In heterozygous Tsix mutants the extent of biallelic Xist expression is reduced, as the loss of Tsix on one allele extents the period between switch-on of the first and second Xist copy [107]. In homozygous Tsix mutants, *cis* repression of Xist is lost on both alleles, speeding up Xist switch-on and increasing the extent of biallelic Xist expression [24]. Since the local double negative feedback has been destroyed, alleles with a *Tsix* deletion have also lost the ability to memorize alternative Xist expression states. The global negative feedback of tXA silencing destabilizes the biallelic Xist expression state, so that cells oscillate between biallelic and no Xist expression.

In fact, our model predicts a similar effect in male *Tsix* deletion mutants: Loss of Tsix on the single X chromosome would allow Xist upregulation, resulting in tXA silencing, which would in turn destabilize Xist expression. Published data on the phenotype of male *Tsix* deletion mutants are contradictory, with some studies reporting faithful Xist repression, and others reporting ectopic Xist upregulation in at least a fraction of cells [20, 22, 174]. Overall, the effects are milder than predicted by the model. Also in our female homozygous *Tsix* deletion mutants we have observed a milder-than-expected phenotype, with the fraction of biallelically Xist expressing cells being only slightly higher than in the WT. Possibly, the effect is alleviated because Tsix transcription has been reduced, but not abolished completely. In line with this, data from our lab profiling nascent transcription suggest that especially during differentiation a considerable portion of Tsix transcription originates from two upstream TSSes within the Xite region[1], and these could still drive Tsix expression in our deletion mutants. One study in fact observed a similar gradual effect in male cells with perturbed Tsix transcription. When comparing mutants

[1] Till Schwämmle, Rutger Gjaltema, Edda Schulz, personal communication

with different levels of residual Tsix expression, they found that Xist was most efficiently upregulated in the mutants with the strongest reduction of Tsix transcription [174]. Two other studies observed ectopic Xist upregulation in XO cells carrying a deletion encompassing both, Xite and Tsix, with a frequency comparable to female cells [175, 176]. In mouse, Tsix' role as a cXR is therefore well established, and experimental data are also consistent with Tsix being involved in mediating a local double negative feedback that allows to memorize alternative Xist expression states. Other (Tsix-independent) positive feedbacks might exist that contribute to stabilizing alternative Xist expression states on Xa and Xi, for instance by establishing differential chromatin modifications at the *Xist* locus on Xa and Xi. They could attenuate the effects observed upon Tsix perturbation.

Tsix transcription so far has however not been detected in cells of other species, making it questionable whether its potential action as cXR is conserved across species. The only exception are human embryonic carcinoma cells, where TSIX has been detected albeit with a shorter overlap to *XIST* [29]. Our model predicts that also this reduced overlap would be sufficient to establish and maintain monoallelic XIST expression.

In our core network cXR thus fulfills many functions. (1) It contributes to the timescale separation between the initial XCI event and the negative feedback (by slowing down Xist upregulation), (2) it implements the ultrasensitivity which generates the threshold response required for the XCI all-or-nothing decision in female versus male cells, and (3) it allows stable maintenance of alternative Xist expression states on Xa and Xi. With a mechanistic model of Xist and Tsix transcription we have shown, that mutual repression between Xist and Tsix can implement all of these functions. However, they could in theory also be mediated by other regulatory principles that could act alone or in concert with Tsix. They will be discussed below.

3.2.1 Alternative molecular implementations of the XCI initiation threshold

In our model, mutual repression between Xist and cXR generates a precise threshold for Xist upregulation, that is only passed in cells with more than one active X chromosome. Such ultrasensitivity in the Xist response could alternatively arise for instance through cooperative binding of transcription factors to *Xist*, which can occur when multiple binding sites are clustered together. A promising candidate region for such cooperative binding is a highly conserved cluster of binding sites within *Xist's* 5' region. After XCI onset the transcription factor Yy1 binds this cluster specifically on the Xist expressing chromosome [177, 178] (Fig 3.4a). Interestingly, two other TFs share this binding motif with Yy1: Yy2 which is encoded on the X chromosome and Rex1, a known Xist repressor that is targeted for degradation by the X-encoded XCI activator Rnf12 [36, 37, 179]. Overexpression of Rex1 decreases Yy1 binding to this site although global Yy1 expression levels are unaffected [178]. In addition, Yy1 knockdown strongly impairs Xist upregulation while Rex1 depletion results in ectopic Xist upregulation [35, 36, 178]. Together, this suggests that Rex1 exerts its function as Xist repressor at least partially through competition for binding with Yy1. Yy1 can thus only bind if Rex1 levels are low and the binding sites are unmethylated (Fig 3.4a) [177, 178]. Upon differentiation, Rnf12 upregulation depletes Rex1

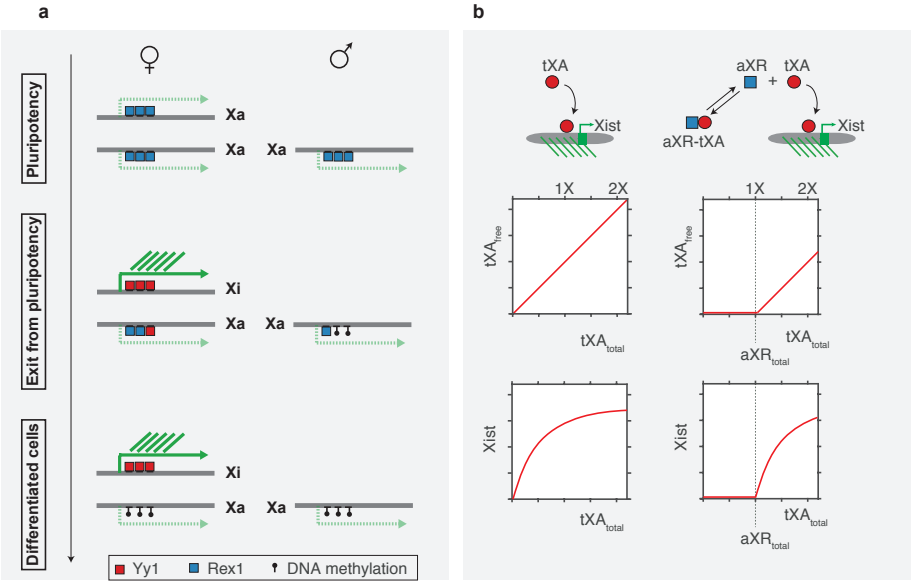

FIGURE 3.4: Possible cXR-independent implementations of ultrasensitive Xist switch-on. a) Cooperative binding of TFs to a binding cluster within *Xist's* 5' region. Competitive binding of Rex1 and Yy1 could ensure that Yy1 can only bind and activate Xist in female cells, where Rex1 levels are lower due to double Rnf12 dose. This allows Yy1 binding to the future Xi, while the Xa and the single X in males remain occupied by Rex1. In differentiated cells, methylation of the binding sites on the Xa (possibly enforced by Tsix transcription) stably prevents Yy1 binding. b) Molecular titration allows ultrasensitivity without cooperativity. An autosomally encoded Xist repressor (aXR, blue) could sequester tXA (red), such that free unbound tXA only exists and can activate Xist in female cells.

in female cells, generating the opportunity for Yy1 binding and Xist upregulation (Fig 3.4a, left) [36, 178]. According to this model, binding sites in males would remain occupied by Rex1 due to lower Rnf12 levels (Fig 3.4a, right). In addition, faster global DNA methylation dynamics in males might contribute to stably preventing Yy1 binding, as the binding sites might get methylated earlier [178]. Upon successful dosage compensation, global DNA methylation levels also increase in female cells, and the binding cluster on the Xa becomes methylated, preventing Yy1 binding (Fig 3.4a, left). Possibly, the methylation of the binding site is promoted specifically on the Xa by active Tsix transcription, which has been shown to promote DNA methylation [27]. The Tsix-mediated feedback and a potential chromatin-based feedback loop might therefore act together and the latter could take over stabilization of the two states upon Tsix downregulation. Yy2 is still poorly characterized, but it could potentially contribute to female specificity if it acts as an Xist activator as it is encoded on the X chromosome: it could tilt the balance against Rex1 binding specifically in female cells. The clustered Rex1/Yy1(/Yy2) binding site could thus generate the threshold required to distinguish between a single and a double dose of Rnf12, which controls the allelic probability of the cluster to be occupied by Rex1.

Yet another mechanism that can produce ultrasensitive responses even in the absence of cooperativity is molecular titration [120, 121]. Here, the threshold does not arise locally at the *Xist* locus but globally. An autosomally encoded Xist repressor could for instance sequester tXA such that free unbound tXA can only exists and activate Xist if expressed in a double dose (Fig 3.4b). Ultrasensitivity arises at the point where almost all inhibitor is bound by tXA, so that any small increase in the total tXA concentration directly translates into unsequestered tXA that can activate Xist. Alternatively, the autosomal repressor could remain unbound and able to exert its function only in male cells. The X-linked Xist activator Jpx has been proposed to

allow Xist upregulation in female cells by evicting Ctcf from the *Xist* locus [42]. Ultrasensitivity could arise if Jpx can sequester all Ctcf in female cells, while two-fold reduced Jpx expression in males would translate into unsequestered Ctcf that can repress Xist. However, Jpx is a rather lowly expressed RNA that would need to sequester the very abundant Ctcf protein. In addition, Ctcf binds to a large number of genomic sites, so it remains unclear how the sequestration could be specific to the *Xist* locus.

Furthermore, the ultrasensitive response is not necessarily implemented at the level of Xist but could also arise if tXA controls the production of yet another potentially autosomally encoded downstream activator in an ultrasensitive manner. This factor should then only be expressed in female cells and therefore does not need to activate Xist in a non-linear manner.

In summary, the required threshold response could be implemented globally, if the activator is only free to exert its function in female cells, for example, through sequestration by an autosomal repressor. Alternatively, the threshold can arise locally at the *Xist* locus for instance through cooperative binding of tXA to the *Xic*, or of course through a local positive or double negative feedback.

3.2.2 Maintenance of monoallelic Xist expression

While monoallelic silencing of the inactive X chromosome in somatic cells has become independent of the Xist RNA, maintenance of the silenced state initially depends on continuous Xist expression [157, 180–182]. It therefore strictly requires maintenance of alternative Xist expression states on Xa and Xi by dedicated *cis*-acting mechanisms that memorize the alternative allelic expression states, such as mutual repression between Xist and Tsix. Alternatively, one of the earliest XCI models proposes stable integration of a DNA element into the active X chromosome [129]. This would constitute an irreversible change that could explain different expression states because the two DNA copies would not be identical anymore. However, no evidence for such DNA sequence changes has emerged yet despite large-scale sequencing efforts. In addition, the insert would have to be precisely removed during meiosis in order to avoid permanent activation or inactiviation of individual X chromosomes. Our observation that initially Xi choice is reversible, as biallelic Xist expression can be reverted to monoallelic expression, provides a final, formal argument against this hypothesis. Thus a self-reinforcing mechanism of some sort seems to stabilize the alternative expression states.

Another biological process where alternative expression states are maintained at two copies of the same gene is genomic imprinting. Here, the choice of the active allele is not random but determined by parental origin. Differential DNA methylation plays an important role in setting up and maintaining genomic imprints [183, 184]. Indeed, DNA methylation has also been shown to be essential to sustain Xist repression on the Xa during the maintenance phase as knockout of the maintenance methyltransferase Dnmt1 results in ectopic Xist upregulation from the Xa [62]. However, DNA methylation is not essential for the establishment of alternative Xist expression patterns, because they are set up correctly in the absence of de novo DNA methylation [185]. Possibly, in the establishment phase DNA methylation acts together with other mechanisms to repress Xist, such as Tsix transcription, which also reinforces DNA methylation

[27]. As cells differentiate, Tsix expression is shut off, and DNA methylation becomes essential for stabilizing the monoallelic expression state. However, there may be alternative Tsix (or cXR)-independent mechanisms that might stabilize the allelic Xist expression states during the establishment phase. Our ODE model analysis has demonstrated that XaXi maintenance can also be achieved by a pure positive instead of a double negative feedback. Epigenetic memory often relies on chromatin-based positive (or double negative) feedbacks, for instance, generated by mutual stimulation of CpG and H3K9 methylation or by mutual antagonism between activating and repressing histone modifications [130, 132, 133]. Alternatively, physical mechanisms like phase separation could also guarantee stable Xa and Xi maintenance. Xist expression has been suggested to seed a protein condensate that excludes RNA PolII and eventually becomes Xist RNA-independent [81].

3.2.3 Emerging considerations on the role of cXR

Regardless of the various alternative implementations of the threshold response and XaXi maintenance, several arguments speak in favor of a double negative feedback that relies on Xist repression by cXR and cXR silencing by Xist. Various experimental perturbations of Xist's silencing ability have resulted in reduced Xist expression level, suggesting that efficient Xist RNA-mediated silencing of X-linked genes is required to reinforce Xist expression: Deletion of the repeat A region, which has been shown to be essential for silencing, results in failure to stably upregulate Xist [12, 73]. Knockdown of components of the m6a methylation machinery, which has also been reported to contribute to Xist's silencing capacity, impairs Xist upregulation, as does deletion of the Xist 5' m6A region [69]. In addition, KO of the silencing factor Spen abolishes Xist upregulation during differentiation of female mESCs, and even reduces Xist levels in inducible systems [68, 74]. This indicates that Xist RNA-initiated, Spen-mediated silencing is required to enforce Xist expression. A recent publication has further narrowed down this autoregulatory effect: The failure of homozygous Spen KO mutants to upregulate Xist upon differentiation can be rescued by a heterozygous *Tsix* deletion [74]. Thus, Xist RNA-mediated silencing of Tsix is required (and possibly even sufficient) for stable Xist expression, which is very much consistent with the notion that Xist expression is stabilized by a double negative feedback of mutual repression between Xist and Tsix.

Taken together, it appears likely that some sort of self-reinforcing feedback generates a memory of the allelic Xist expression state, and experimental data are consistent with the idea that this autoregulatory feedback is mediated by mutual repression between Xist and Tsix. In addition, female-specific induction of XCI requires a threshold response that is either implemented at the level of Xist or an upstream regulator.

3.3 Role of the tXA - dosage sensing and mutual exclusivity

The tXA implements the global negative feedback that couples the two X chromosomes and moves the cell from the regime where Xist upregulation is triggered (2x tXA, XaXa) into the

bistable regime where Xa and Xi Xist expression states are stable (1x tXA, XaXi). It contributes to mutually exclusive choice and generates a safety net for cells that accidentally upregulate Xist from both alleles.

To ensure female specificity the tXA must be expressed at two-fold higher levels in females than in males and it must activate XCI in a switch-like manner within this two-fold range. To ensure a reliable distinction between expression levels in males and females, tXA must likely be a rather highly expressed gene to reduce noise induced fluctuations in its expression level. To mediate the required fast-acting negative feedback, tXA must be rapidly silenced during XCI. Overexpression of tXA in males should induce ectopic XCI, while its heterozygous deletion in females should abolish XCI. The only known tXA candidate so far is the E3-ubiquitin ligase Rnf12/Rlim, which is encoded in *Xist's* proximity within the *Xic* [35]. Rnf12 targets the autosomally encoded Xist repressor Rex1 for degradation, and thereby activates Xist [36, 37]. Rnf12 clearly is a dose-dependent Xist activator as its overexpression in male cells results in ectopic XCI [35]. Its deletion in females has different effects depending on the precise culture conditions. In vivo, its complete absence prevents imprinted XCI in females, but random XCI has been reported to be unaffected [39, 40]. In vitro, complete Rnf12 absence abolishes random XCI in some culture conditions, but not in others [38, 39, 41]. Its heterozygous deletion only delays but does not abolish random XCI [35]. The asymmetric effects caused by overexpression and deletion of Rnf12 indicate that additional tXA factors exist. If multiple tXA factors cooperate to ensure XCI activation, overexpressing a single one in males should indeed cause ectopic Xist upregulation, while its (heterozygous) deletion in females does not necessarily abolish XCI as the other tXA factors might compensate its loss.

It is difficult to distinguish between additive or redundant action on the basis of these perturbation phenotypes. If multiple tXA factors act in a redundant manner through completely independent pathways, one would expect random XCI to be unaffected by Rnf12 deletion. However, the precise environmental conditions might determine whether the other factors are expressed and can exert their function, possibly explaining why Rnf12 KO abolishes random XCI in some culture conditions but not in others [38, 41]. If multiple tXAs act together in an additive manner through the same pathway, one might expect a milder phenotype from heterozygous than from homozygous Rnf12 deletions, as the other factors can compensate a two-fold reduction more easily than a complete loss. Whatever its exact implementation, the cooperation between multiple tXA factors might contribute to robustness of X-dosage sensing. Multiple tXA factors can more reliably average out gene expression noise, which in single cells frequently extends beyond the two-fold expression difference between males and females [186].

An X-linked XCI activator that could potentially cooperate with Rnf12 in Xist activation is the lncRNA Jpx, which is also encoded in the *Xic* [43, 187, 188]. As discussed above, Jpx has been proposed to activate Xist by sequestering away Ctcf from the *Xist* locus [42]. Besides the already discussed problems with this titration hypothesis, it also still remains controversial whether Jpx actually acts in *cis* or *trans* [42–44, 188]. In addition, Jpx is one of the few genes that escape XCI and therefore cannot mediate the required global negative feedback. Another XCI-activating lncRNA within the *Xic* is Ftx, but Ftx acts in *cis* by the act of transcription and can therefore not cooperate with Rnf12 in *trans*-activation of XCI [45, 189]. Therefore additional

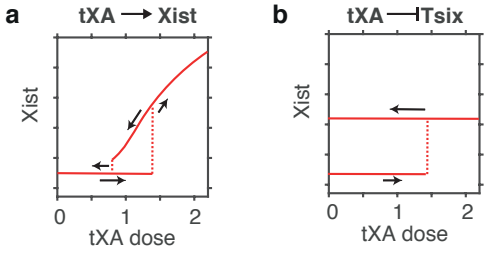

FIGURE 3.5: Hysteresis and irreversibility [123]. a) tXA activates Xist: The system exhibits a different response depending on whether it was previously in its "on" or "off" state (Hysteresis). This allows Xist to stay "off" on the Xa, while being "on" on the Xi. b) tXA represses Tsix: The system might stay in its "on" state indefinitely even after the tXA trigger is completely removed (Irreversibility). This would prevent reversal of erroneuos biallelic Xist expression.

tXA factors must exist and remain to be identified.

There is solid experimental evidence for the existence of a (silencing-mediated) negative feedback: We demonstrate that cells are able to sense erroneous upregulation of Xist from both alleles and to correct it to monoallelic expression. A very elegant experiment by Barakat and colleagues also suggests that at least one active Rnf12 allele is required, not only to initiate but also to maintain random XCI: If *Rnf12* and *Xist* are deleted on the same X chromosome, this will always result in inactivation of the WT X, because it carries the only intact Xist allele, and therefore in silencing of the only intact Rnf12 copy. These mutant mESCs are unable to maintain XCI indicating that continuous Rnf12 expression is required for XCI [43].

Lastly, tXA could in principle also act indirectly, for instance by mediating global dosage effects, such as lower DNA methylation levels in female cells reflected in hypomethylation of the *Xist* promoter, which could facilitate Xist upregulation [56, 58]. Alternatively, tXA action could also be implemented through repression of cXR, rather than through direct Xist activation. This could explain why Tsix is expressed at higher levels in undifferentiated XO than XX cells and why Xist from a doxycycline inducible allele can be upregulated more easily in undifferentiated XX than XO cells[2]. Such indirect action of tXA, might however complicate reversal of biallelic Xist expression. If the rate of Xist production does not dependent on the tXA dose, Xist expression might be sustained, even upon complete tXA silencing (e.g. because Xist RNA-mediated silencing of Tsix is very stable, once it has been established). In this case, the allelic memory of the Xist expression state would be so strong, that Xist switch-on would become irreversible because zero tXA stimulus would be required to maintain the locus in the flipped state (Fig 3.5b). During XCI onset, the hysteresis must be strong enough to allow maintenance of the Xi upon reduction of the tXA stimulus to a single dose, but it should not be strong enough to make Xist expression self-perpetuating in the complete absence of tXA, which would make the Xist switch-on irreversible and prevent cells from correcting erroneous biallelic Xist expression (compare Fig 3.5a and b). Dosage sensing may thus be implemented by multiple tXA factors, with at least one of them acting directly on Xist, to ensure reversal of biallelic Xist upregulation, and the other factor(s) indirectly activating Xist, for instance through repression of Tsix or by mediating global changes that facilitate Xist upregulation in females.

[2]Ilona Dunkel, Till Schwämmle, Rutger Gjaltema, Edda Schulz, personal communication

3.4 cXR-tXA - a unifying core network of XCI in different species

Our two-regulator network can generate alternative routes to the monoallelic expression state. Depending on the precise parameter regime, cells either directly upregulate Xist in a monoallelic fashion or transit through a period of biallelic Xist expression which is then resolved to monoallelic expression. The capacity of the model to generate different extents of transient biallelic Xist expression is exciting as the frequency of biallelic Xist upregulation differs between species. It is less frequent in mice than in rabbit, but in both species it is quickly resolved to monoallelic expression [99, 141, 165]. In human embryos, Xist is initially upregulated from both X chromosomes in all cells but fails to induce complete silencing and persists for several days before it is resolved to monoallelic expression [99, 100]. In our model, the relative timescales of Xist switch-on and tXA silencing determine the extent of transient biallelic Xist expression. The higher biallelic level in rabbits and humans compared to mice could thus be the result of faster Xist switch-on and/or slower tXA silencing. Recent work has found that overall protein stability is increased more than two-fold in human compared to mouse which can explain differences in the speed of development between these species [190, 191]. An increased half-life of the tXA factor in human and rabbit compared to mouse, would in fact result in a longer silencing delay, and would therefore be predicted to result in a higher extent of biallelic Xist expression.

Since Tsix transcription has not yet been detected in human or rabbit embryos it is questionable whether Tsix also fulfills the role of cXR in these species. We however showed, that in theory the *XIST/TSIX* locus could also generate robust monoallelic Xist upregulation with a reduced overlap as it has been reported in human embryonic carcinoma cells. Besides TSIX, an alternative promising cXR candidate in human is the lncRNA XACT, which co-accumulates with XIST on both Xas in human preimplantation embryos and naive hESCs, but is specifically expressed from and coats the Xa in post-XCI primed hESCs, where XIST coats the Xi [30, 31]. This mutually exclusive expression pattern is consistent with a double negative feedback of mutual inhibition that memorizes the alternative XIST expression states. We have also adressed the human-specific extended biallelic XIST expression period during preimplantation development, and find that our model can reproduce such extended biallelic expression by making one of two alternative assumptions: Either (1) cXR is only upregulated once the biallelic expression state is resolved, or (2) it is dampened by XIST expression during the biallelic expression period, while tXA is unaffected. Hypothesis 1 is interesting because it might explain why TSIX has not been detected in human pre-XCI naive ESCs (or human preimplantation embryos). On the other hand this hypothesis is incompatible with a role of XACT as cXR because XACT is clearly co-expressed with XIST from both Xas prior to XCI. The recent establishment of a model system that allows researchers to observe the onset of random XCI in hESC will hopefully soon shed further light on this matter, as it will allow time-resolved expression measurements and perturbations of cXR candidates [101].

To learn more about regulator conservation between species, we might also turn to hybrids such as hinny and mule. Here, all X-linked regulators are present in one copy from the horse and one copy from the donkey. Interestingly, female hinnies and mules have been described to

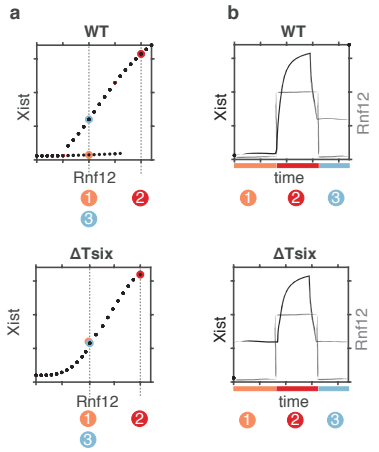

FIGURE 3.6: Experimental assay to measure memory at the *Xist* locus. On an allele in the Xist^{low} state (orange) transient overexpression of Rnf12 should induce Xist expression (red), the positive feedback should then stabilize the Xist^{high} state even upon subsequent reduction of Rnf12 levels (blue). a) Predicted allelic Xist steady state for different Rnf12 doses in WT (top) and ΔTsix (bottom). b) Timecourse of Rnf12 overexpression and resulting Xist expression levels in WT (top) and ΔTsix (bottom).

undergo random XCI [192, 193]. This indicates that (1) monoallelic Xist upregulation is still functional, thus that *trans* regulators seem to be conserved, and (2) choice of the Xi is random, thus (*cis*) regulators controlling Xist switch-on time are conserved. Another interesting indication about regulator conservation comes from studies which have inserted human *Xic* transgenes into male mouse ESCs and find that they can ectopically induce mouse XCI, suggesting that (at least) *trans*-acting factors are conserved between mouse and human [194, 195]. Curiously, human XIST does not induce efficient gene silencing and is already upregulated before differentiation is induced. The latter could indicate that cXR indeed is only upregulated later in humans.

Thus while the identity of the regulators, or even the precise implementation of the two feedbacks might vary to different extents between mammals, the network architecture may well be conserved.

3.5 Outlook

In summary, a global tXA-mediated negative feedback combined with a local cXR-mediated toggle switch can ensure female-specific and monoallelic XCI. This work provides experimental evidence for the existence of the global negative feedback through the demonstration that cells are able to reverse biallelic Xist expression. Rnf12 clearly acts as a dose-dependent Xist activator but additional factors must exist. These factors could be identified by analyzing their ability to induce XCI when they are overexpressed in male cells rather than assessing their deletion phenotype in female cells where effects might be masked by cooperating factors.

To shed light on the exact implementation of the threshold response and the positive feedback, one should try to measure both experimentally. An essential factor that limits efforts in this direction is that the identity of the tXA is still not fully uncovered. However, with Rnf12 a regulator is known that has been shown to activate Xist in a dose-dependent manner. Rnf12 could be titrated to analyze whether Xist responds in a non-linear manner to changes in the tXA

dose. This should allow the observation of the ultrasensitive response independent on whether it is implemented at the level of Xist or an upstream regulator but of course only if Rnf12 is involved in the switch-like response, e.g. by controlling Rex1 levels and therefore binding probability to the *Xist* 5' binding cluster. Once the ultrasensitive response can be captured, candidate mechanisms could be perturbed to test their role in its implementation. For instance it could be assessed whether titration of Rnf12 still results in a non-linear Xist response if the *Xist* 5' binding cluster is deleted, if *Tsix* is deleted, or if both are deleted together. To measure the allelic memory, Rnf12 should be transiently overexpressed to a level beyond the threshold to induce Xist upregulation from an allele that was previously in the $Xist^{low}$ state (Fig 3.6). The memory-generating mechanism should now stabilize Xist expression even upon subsequent reduction of Rnf12 levels. Again, perturbation of the memory-mediating mechanism(s) (such as Tsix transcription) should destroy the memory (Fig 3.6, compare top to bottom). If multiple mechanisms act together in a redundant manner, they must be perturbed in combination. Besides the hysteresis, another important feature of any bistable system is its ability to toggle between the two alternative states. By giving a trigger stimulus that impacts the double negative feedback, one should thus be able to flip the allelic Xist expression state between ON and OFF. One could for instance investigate whether transient stimulation of Tsix on the Xi can stably silence Xist. The notion that antisense transcription might be capable of mediating toggle switches that can maintain alternative transcription states is also worth further investigation, which is what the second part of this thesis sets out to do.

Part II

Transcriptional memory at antisense pairs

4 Introduction

In the first part of this thesis we have demonstrated that antisense transcription could stabilize alternative expression states of Xist on the Xa and the Xi and might thereby generate long-lasting transcriptional memory. Such antisense transcription is not limited to the Xic but is wide-spread across genomes in all three domains of life [196–201] and is found for both, genes with and without coding potential [202, 203]. In humans it is estimated that ∼20-40% of transcribed regions have an antisense transcript [204–207], and some of these transcripts are well conserved across species [208, 209], indicating a functional role. It is thus tempting to speculate that antisense transcription could be involved in maintaining alternative transcription states also in other contexts.

In the second part of this thesis, we set out to understand whether the generation of transcriptional memory could be a widespread function of antisense transcription. By "transcriptional memory", we mean the ability of an antisense locus to retain the transcription state that it has acquired due to some past transient signal, generating *cis*-encoded memory. To investigate this, we build a mathematical model to probe for transcriptional memory by exposing an antisense locus to different initial transcription states (sense ON & antisense OFF, or sense OFF & antisense ON) and then analyzing its potential to maintain these alternative transcription states stably.

At the *Xist/Tsix* locus memory arises because the two antisense transcription units mutually repress one another by TI and RNAP- or RNA-mediated promoter silencing. The ability of the Xist RNA to silence its antisense gene *Tsix* (and other X-linked genes) is a peculiarity, mediated by the Xist RNA's repeat regions that recruit silencing factors. Nevertheless, any two convergent genes carry an immense potential to repress one another through transcriptional interference, or post-transcriptionally through RNA:RNA interactions.

4.1 Eukaryotic transcription

To model antisense transcription, we have to consider the reactions that govern transcription. Eukaryotic transcription can be divided into three distinct stages: Initiation, elongation and termination (Fig 4.1) [210]. Transcription usually initiates at a defined position, the transcription start site (TSS), which lies within a core promoter. The promoter provides a binding platform for the transcriptional machinery consisting of RNA polymerase (RNAP) and general transcription factors and thereby controls transcription initiation. The transcriptional activity of promoters is

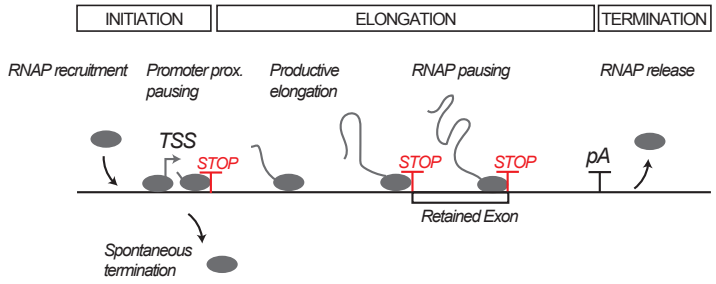

FIGURE 4.1: Distinct phases of the transcription cycle: initiation, elongation and termination. RNAPs are grey ovals (with nascent transcripts). TSS: Transcription start site; pA: polyadenylation site. Figure adapted from [210].

regulated by local chromatin context and (distally located) regulatory elements, called enhancers (for a review see [211]).

Promoters in addition seem to transit between inactive and active states, causing transcription to occur in short intense bursts with several initiation events during the active periods, separated by inactive periods without transcription initiation [151, 212, 213]. Transcriptional activation can thus either occur through an increase in the frequency of these bursts, or an increase in their size (=the number of transcripts that are produced per burst). Measurements of transcription burst frequencies and sizes suggest that while the core promoter elements seem to determine the burst size (k_{off}/k_{ini}), enhancers seem to control the burst frequency (=k_{on}) [214–216]. During initiation, RNAPs are thus recruited to the DNA template, and this recruitment appears to occur in discontinuous bursts.

After RNAPs have initiated, eukaryotic transcription remains a discontinuous process with phases of pausing and phases of productive elongation. Before the RNAPs transit to productive elongation they can be halted in the promoter region for a while, a process termed promoter proximal pausing [217, 218]. Pausing is reflected in measurements of population-averaged RNAP density across the gene body, which reveal a peak of RNAP accumulation at promoters (e.g. ChIP-Seq, NET-Seq) [217]. Another factor that can contribute to the accumulation of RNAPs at promoters is imperfect processivity where a fraction of the promoter-bound RNAPs never goes into productive elongation but rather terminates prematurely close to the transcription start site (TSS) [219, 220]. RNAP firing from a promoter can thus be approximated as a two-step process: (i) binding of the RNAP to the promoter, where they form an open "sitting duck" complex, and (ii) formation of an elongation complex that can be released into productive elongation. Both steps determine the promoter strength as they together dictate the intervals in which a promoter fires RNAPs.

4.2 Transcriptional interference

Simultaneous transcription of two convergent genes imposes a problem: How to deal with two-way traffic on the one-way DNA road? Transcriptional interference (TI) has been defined by Shearwin and colleagues as the "*suppressive influence of one transcriptional process, directly and in cis on a second transcriptional process*" [221]. Wherever two transcription units overlap partially they can interfere with each other's transcription, whether their promoters are oriented

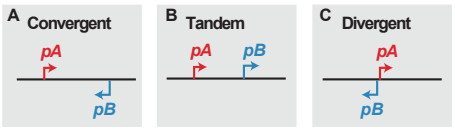

FIGURE 4.2: Promoter arrangements that can result in transcriptional interference. a) convergent promoters, b) tandem promoters, c) divergent promoters.

convergently, with both transcripts partially or fully overlapping, or in tandem, with one promoter lying upstream of the other (Fig 4.2).

In the following section, I will give an overview of the mechanisms of transcriptional interference and the experimental evidence supporting them. As proposed by Shearwin et al., we will focus on TI that is caused by RNAPs, and neglect roadblock interference caused by DNA-bound proteins, as these are likely DNA sequence-specific interactions and we seek to investigate a generic mechanism, that could apply to all convergent promoter pairs.

RNAP collisions

Collisions could occur between two elongating RNAPs that move in opposite directions, and might result in the termination of one or both transcription processes (Fig 4.3a). They were initially proposed by Ward and Murray to explain the interference observed in vivo between the convergently oriented promoters pR and pRE of coliphage lambda [152]. By putting the *GAL7* and *GAL10* genes in *S. cerevisiae* in a convergent orientation, the Proudfoot group showed that in the absence of an intergenic region with transcription termination signals (pAs), the amounts of full-length transcripts produced from both promoters were reduced, although transcription initiation remained unchanged [222]. This suggested the presence of an interfering mechanism during the elongation phase that results in termination of one or both transcriptional processes. Atomic force microscopy has provided the first direct experimental evidence for *E. coli* RNAP collisions at single molecule resolution [153], but the technique cannot investigate whether in vivo the RNAPs dislodge from the DNA template upon collision. In eukaryotes, an experimental setup of two convergent inducible promoters has revealed that two collided RNAPs stay attached to the DNA until they are targeted for proteasomal degradation. Before they are removed from the DNA template they could even generate a transcription block for following RNAPs [154]. In vitro, RNAP collisions thus result in RNAP stalling. In vivo, stalled RNAPs are remarkably stable and are removed by ubiquitylation-mediated degradation. It is still unclear how frequently collisions occur at individual genes, and whether mechanisms might be in place that facilitate bypassing of RNAPs in vivo [154].

Mathematical modeling has proven a useful tool in dissecting the relative contributions of different mechanisms to the total interference felt by a promoter. A seminal model by Sneppen and colleagues found that the interference by collisions increases with increasing activity of the two promoters and increasing distance between them [223], a prediction which has been confirmed experimentally [203]. Interestingly, in human and mouse, the expression levels of natural antisense transcripts seem to correlate negatively with the overlap between sense and antisense transcript [224], concordant with the hypothesis that transcriptional interference becomes stronger for longer overlaps.

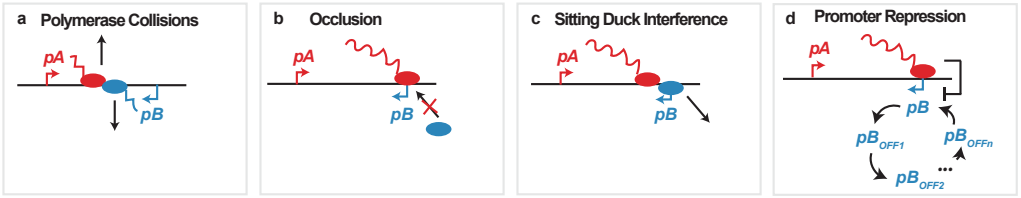

FIGURE 4.3: Mechanisms of transcriptional interference. A pair of convergent promoters pA and pB can interfere with each other's transcription by a) polymerase collisions, b) occlusion, c) sitting duck interference, and d) promoter repression.

Occlusion

Occlusion describes the blocking of the promoter region by antisense RNAPs, such that no new sense RNAPs can initiate transcription (Fig 4.3b). Originally this mechanism was proposed by Adhya and Gottesman to explain the interference that an upstream promoter exerted upon a downstream promoter in a tandem orientation: For the *gal* promoter in the galactose operon of *E. coli* the authors had observed a strong reduction in its transcriptional activity if the upstream prophage promoter *pL* was active [225]. For the above-discussed *Gal7* and *Gal10* genes a similar effect was described: In their native arrangement where the *Gal10* gene lies upstream of the *Gal7* gene, the *Gal7* promoter is strongly reduced in its activity if the upstream *Gal10* gene is transcribed [226]. DNase footprinting in *E. coli* has revealed that for overlapping promoters initiation of transcription cannot occur simultaneously on both promoters [221, 227, 228]. Thus DNA-bound RNAPs seem to block access to the promoter for newly initiating RNAPs.

These experiments are however not always clearly interpretable. Interference that is caused by transcription through a promoter can be a sum of different mechanisms, whose individual contributions can be hard to dissect experimentally. Some of the above-described effects might also arise from sitting duck interference which is discussed below. Modeling can help to dissect these effects. It predicts that interference by occlusion is minor unless the interfering promoter is extremely strong, because the transit time across the promoter is usually shorter than the intervals in which standard promoters fire. Of course, the transit time also depends on the size of the promoter region and the speed of transcription across the promoter, such that for instance pausing of RNAPs over the promoter region could increase the interference by occlusion [221, 223].

Sitting duck interference

Sitting duck interference (SDI) describes the removal of promoter-bound RNAPs, so called sitting duck complexes (SDCs), from DNA upon their encounter with an elongating antisense RNAP (Fig 4.3c). The SDCs are waiting to fire and are not yet in the productive elongation conformation, and therefore likely have a weaker association to the DNA template. SDI was proposed by Callen et al. in 2004 based on observations in the locus that controls the lysis/lysogeny decision in bacteriophage 186. Here the lytic and lysogenic promoter are oriented convergently.

The authors observed that the lytic promoter strongly represses the much weaker lysogenic promoter. This interference was dramatically reduced if transcription from the strong promoter was terminated before it reached the weak promoter, indicating that converging RNAPs must pass the weak promoter in order to repress it efficiently. The interference was also too strong to be caused by occlusion only, which is limited by the fraction of time that the promoter is occupied by an antisense RNAP [203]. In support of the SDI hypothesis, the authors found that a promoter with a reduced ability to form sitting duck complexes was less sensitive to interference. The strength of SDI strongly depends on the retention time of SDCs at the promoter: The sensitivity of a gene towards SDI depends on the kinetic properties of its promoter. Mathematical modeling predicts that SDI is maximal when binding and firing rates are equal: If firing is faster than binding, RNAPs spend less time at the promoter, while if binding is faster than firing, every dislodged SDC is rapidly replaced by another. The model also shows that SDI becomes stronger the more dominant the aggressive promoter is over the sensitive promoter [223].

Promoter repression

There is ample evidence that the act of transcription can also change surrounding chromatin and thereby influence transcription of neighboring genes [172, 205]. Often these effects are activating as they lead to opening of chromatin and repositioning of DNA into the active compartment [229, 230] or cause DNA demethylation [231], but antisense transcription frequently also establishes a repressive chromatin environment at the sense gene (Fig 4.3d). How are these changes established? PolII as well as PolII-interacting factors have been shown to recruit a number of chromatin-modifying enzymes during elongation, which then modify the epigenetic environment around the transcribed locus (see [232] for a review). Transcriptional elongation promotes the establishment of a repressive chromatin context through the recruitment of enzymes that catalyze H3K36me3 which in turn promotes DNA methylation [233–236]. Transcriptional elongation thereby suppresses RNAP initiation within actively transcribed genes.

However, the antisense transcripts themselves can also trigger epigenetic changes, for instance by binding to the complementary DNA strand and recruiting DNA methylating enzymes [237–241] or by acting as a scaffold for the recruitment of histone-modifying enzymes [242, 243]. Although this recruitment likely depends on the RNA sequence or structure, it still appears to be a very prevalent mechanism: The polycomb repressive complex 2 (PRC2) which promotes heterochromatization by histone methylation, for instance, is known to interact with more than 3,000 antisense transcripts [202, 244]. Many other chromatin-modifying enzymes also have RNA-binding motifs, indicating that they could be recruited by (*cis*-acting and) potentially weakly expressed RNAs to locally modify the chromatin environment [245].

In several contexts, antisense transcription thus suppresses sense transcription by the recruitment of epigenetic changes. A few concrete examples are the following: In the mouse model for a genetic form of anemia, where the α-globin 2 (HBA2) gene acquires a ubiquitously expressed antisense RNA, this antisense RNA induces DNA methylation to repress HBA2 transcription [237]. Many imprinted gene clusters are silenced by antisense lncRNAs, for instance through the recruitment of histone and DNA methyltransferases that establish a repressive chromatin environment at the promoter of the imprinted gene [239, 246]. The antisense non-coding RNA

in the *INK4* locus (ANRIL) is overexpressed in a variety of human cancers and silences the tumor suppressor locus *INK4b-ARF-INK4a* through the recruitment of the polycomb machinery that establishes H3K27me3 [247, 248]. In plants, regulation of the *FLOWERING LOCUS C (FLC)*, which encodes a floral repressor, is governed by antisense transcription. Upon prolonged exposure to cold, FLC's antisense transcript COOLAIR is upregulated and silences FLC to allow flowering after winter cold [249, 250]. Last but not least, Tsix has been shown to repress Xist by the establishment of repressive histone modifications at the Xist promoter [28].

Thus, convergently oriented genes with overlapping gene bodies can potently suppress transcription initiation from the antisense promoter. The act of transcription itself can establish a chromatin context that represses transcription initiation (e.g. H3K36me3 and DNA methylation), but also the transcripts themselves can recruit enzymes that catalyze repressive chromatin marks (e.g. PRC2, histone methyltransferases).

4.3 RNA-mediated interactions

Two overlapping genes transcribed from convergent promoters can produce perfectly complementary transcripts that have the potential to form duplexes. These RNA:RNA interactions can regulate the expression of the sense transcript in multiple ways: They can reduce translation efficiency by preventing ribosome loading [251, 252]. They can increase RNA stability by masking sites that induce degradation (e.g. miRNA binding sites [240, 253]), or decrease stability through targeting by RNases that only degrade dsRNA [254]. They can even induce transcription attenuation by the formation of a terminator structure that prevents RNAPs from transcribing further [255–257]. Such interactions are however context dependent (splicing, RNA secondary structure). We have therefore not included them in our analysis.

Thus, convergent genes can negatively regulate one another in multiple ways, potentially allowing non-linear responses to regulatory input, as antisense transcription can filter out low level sense transcription. Mutual repression between a pair of antisense genes might even allow each strand to self-reinforce its own expression by repressing its convergent transcription unit, possibly allowing memory to arise.

4.4 Aim of the study

Being able to make decisions and stably maintain them is an essential part of multicellular life as different cell fates must be encoded by identical DNA sequences. During cell fate decisions a transient stimulus triggers differentiation into a specific fate but this fate is maintained long after the stimulus is gone. Oftentimes this requires *trans* memory: The cell as a whole makes a decision and must remember this decision. Transcriptional interference however acts in *cis* only, and might therefore even be capable of generating a *cis* memory, where two copies of DNA within the same nucleus assume and maintain different transcription states, such as Xist during XCI.

We seek to systematically analyze the potential of transcriptional interference between antisense genes to generate *cis* transcriptional memory. We deliberately focus on analyzing the effects of transcriptional interference, although any two antisense genes might also regulate each other by RNA:RNA interactions. Transcriptional interference, however, is an inherent property of all antisense loci, and therefore likely a generic mechanism by which antisense transcripts exert their function. It is tempting to speculate that the establishment of transcriptional memory could be a widespread function of antisense transcription. Understanding the factors that control transcriptional stability might in addition pave the way for using antisense transcription units to achieve desired behaviors in synthetic gene circuits.

5 Results

5.1 Antisense transcription can generate transcriptional memory

We set out to understand whether the act of antisense transcription alone, in the absence of RNA-mediated interactions, is able to stably maintain alternative transcription states. To this end, we developed a mathematical model to simulate transcription at an antisense locus (Fig 5.1). We probed for transcriptional memory by exposing the locus to different intial transcription states and analyzing its potential to maintain these stably. To identify the model reactions that are critical for the establishment of such transcriptional memory, we then simplified the model and analyzed the maximally simplified model version with respect to its parameter values.

We simulated a symmetric locus architecture with both genes completely overlapping one another, enabling mutual promoter repression (Fig 5.1). The distance between the two promoters (=the overlap between the two genes) was varied between 5 kb and 50 kb ($L_{Overlap}$). To model transcription initiaton we included promoter-proximal pausing and premature termination of promoter-bound polymerases since experimental measurements indicate an increased density of polymerases in the promoter-proximal region [217–220]. For simplicity we assumed that pausing would directly occur at the promoter segment, a justifiable simplification, as measurements in mammalian cells indicate that pausing occurs on average 50 bp downstream of the TSS [210, 258–260]. In the model, promoter-bound RNAPs are halted at the promoter for a specific time, determined by the release rate $k_{release}$. Upon their release they either go into productive elongation, or they fall off the gene with probability p_{term} (Fig 5.1). Such multi-step initiation affects the retention times of RNAPs at the promoter and can thereby influence the strength of interference felt by a promoter.

In our model, transcriptional interference occurs by four distinct mechanisms (Fig 5.1):
(i) RNA polymerase collisions occur between sense- and antisense transcribing polymerases with one randomly chosen polymerase being removed from the gene. (ii) Antisense polymerases transcribing through the sense promoter region induce a transition to the OFF state (promoter repression, PR). Measurements of transcription from endogeneous mammalian genes indicate that silent promoter periods show a peaked distribution suggesting a multi-step progression towards transcription activation [261–263]. To approximate this, we modeled promoter transitions as an irreversible cycle of one active state (ON) where transcription can occur and n sequential inactive states (OFF) [264, 265]. We assume that the mean time spent per OFF state is the same

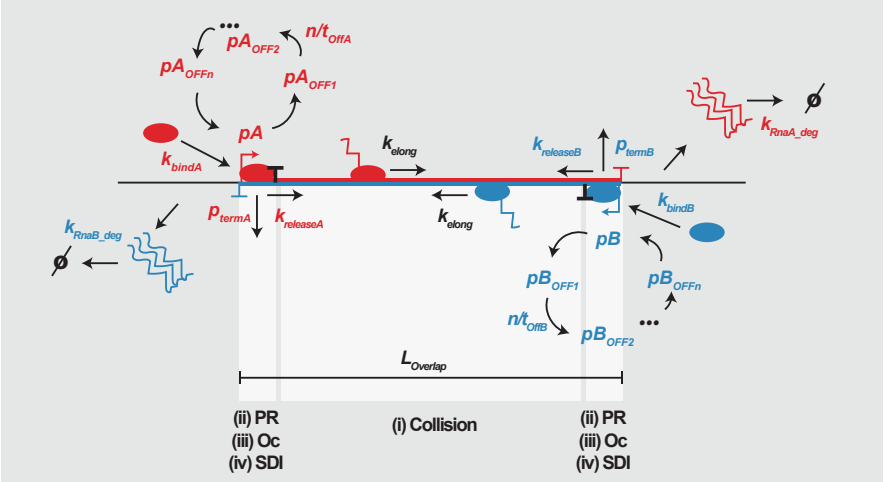

FIGURE 5.1: Schematic representation of the antisense model to investigate stability of alternative transcription states of two convergent genes A and B (A high and B low OR A low and B high). The model describes transcription initiation, transcription elongation, and RNA degradation. The promoter is turned off by passing antisense polymerases and the OFF state has a lifetime determined by t_{OffA} or t_{OffB}, respectively. Transcription of A and B mutually interfere by collisions, promoter repression (PR), occlusion (Oc) and sitting duck interference (SDI).

for each inactive state, such that the rates of transition to the next state are equal and determined as the number of OFF states divided by the mean total silent period ($k_{rev-rep} = \frac{n_{Off}}{t_{Off}}$). This implementation allows us to vary the duration and number of the rate-limiting steps that give rise to the refractory periods [264]. (iii) Occlusion prevents recruitment of new RNAPs to the promoter, if the promoter region is occupied by an antisense polymerase. (iv) Upon an encounter between a promoter-bound RNAP and an elongating antisense RNAP, we assume that the promoter-bound RNAP always dislodges given the structural differences between SDCs and ECs.

The reactions describing transcription initiation and elongation of the two antisense genes were combined into a mathematical model, and individual alleles were simulated using the stochastic Gillespie algorithm [136]. In brief, RNAPs can bind the promoter and while bound at the promoter interfere with antisense RNAPs by SDI. Once they are released into productive elongation, they interfere with antisense RNAPs by collision. While passing their antisense promoter they prevent binding of new antisense RNAPs to the promoter by occlusion, and they can also induce a transition of the antisense promoter to the inactive state (Fig 5.1). The DNA was divided into segments of 100 bp, as a broad estimate of the footprint of an RNAP on the DNA template [219]. In the simulation, elongation occurs at fixed time intervals (2.5sec/100bp), inferred from measurements of RNAP speed ([149], Table 5.1). Between elongation steps, all other reactions are simulated with the stochastic Gillespie algorithm to account for random fluctuations.

Different values of the parameters describing transcription initiation and duration and structure of the silent promoter periods were systematically tested. Table 5.1 summarizes the parameters of the antisense model. We varied the total lifetime of the silent state between one minute and 10 hours, and the number of promoter OFF states between 1 and 5 since refractory periods of most mammalian promoters are well described by five sequential OFF states and last roughly one hour [261, 262, 264]. For the rate of RNAP binding to the promoter and release from the

promoter values between 10 to 1000 molecules/h were tested, and the probability for pre-mature termination instead of release into productive elongation was varied between 5% and 90%.

Table 5.1: Parameter values of the general antisense model

Description	Parameter	Parameter Values
Elongation rate [bp/sec]	k_{elong}	40 ([149])
Binding rate of RNAPs to pA or pB [h^{-1}]	k_{bindA} or k_{bindB}	10 - 1000 (logarithmically distributed) ([266])
Release rate of RNAPs from pA or pB [h^{-1}]	$k_{releaseA}$ or $k_{releaseB}$	10 - 1000 (logarithmically distributed) ([266], [267])
Average time of pA or pB in OFF state [h]	t_{OFF-A} or t_{OFF-B}	1/60 - 10 (logarithmically distributed) ([264], [266])
Probability for pA- or pB-bound RNAP to terminate [1]	p_{termA} or p_{termB}	0.05 - 0.9 (logarithmically distributed) ([267], [220])
No. promoter OFF states [1]	n	1 -5 (linearly distributed) ([264])
Overlap between A and B [kb]	$L_{Overlap}$	5 -50 (linearly distributed)

To investigate whether transcriptional interference allows the antisense locus to maintain a memory of its transcription state, we simulated the model with >275,000 parameter sets randomly sampled from the complete parameter space. To probe for transcriptional memory, two simulations with different initial conditions were performed for each parameter set: In the first one, gene A is transcribed at the steady state level in the absence of transcriptional interference, while gene B is not transcribed and its promoter is repressed. In the second one, gene B starts transcribed and gene A is repressed. Transcription was simulated over a period of 500 h and the simulation was repeated 100 times per parameter set.

First, we set out to test whether the locus can retain a memory of its initial state. As a measure for the stability of the transcription state we calculated the first time point at which the normalized ratio of polymerases on A and B would cross 1 (see section 9.3.2.1). This is the first timepoint at which the gene that initially started in the OFF state feels less repression by its convergent promoter than the gene that started in the transcribed state (Fig 5.2a). We termed this measure the first-switching-time (fsw). Fig 5.2a shows an example simulation from both initial conditions with the first-switching-times (fsw) indicated. We calculated the average first-switching-time across all simulated alleles for each parameter set. Parameter sets that are able to maintain alternative transcription states should have a high average first-switching-time in both initial conditions, maximizing the minimal average first-switching-time across the two initial conditions (Min(fsw1,fsw2), MFS). Fig 5.2c shows example simulations for three sets that were classified as unstable (left, MFS<2h), medium stable (middle, 20h<MFS<100h) and highly stable (right, MFS>200h). In the unstable example set, gene B immediately gains dominance, independent of the initial condition. In the stable set, gene A remains dominant for a long time with initial condition 1, while B remains dominant with initial condition 2. In the category with medium stability, the first switch does not occur immediately but after a few hours. For $\sim 0.5\%$ of parameter sets the two alternative transcription states were stably maintained for >200 h (MFS>200 h, Fig 5.2b). In principle, transcriptional interference between two antisense genes can thus maintain alternative transcription states for several hours to days.

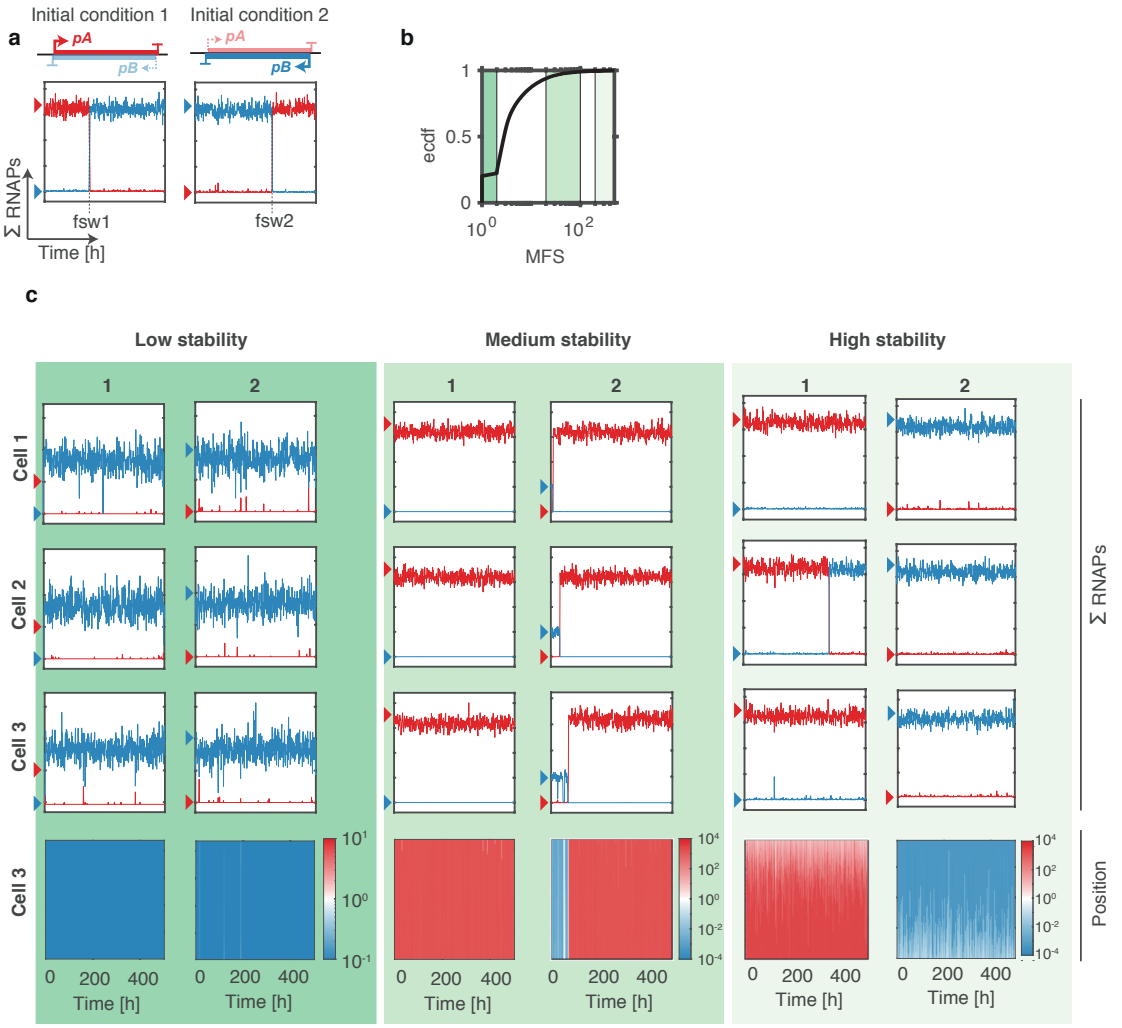

FIGURE 5.2: Minimal first-switching-time (MFS) as a measure for transcriptional stability. a) Example simulation of one parameter set with both initial conditions (red: gene A, blue: gene B). Colored arrows indicate initial conditions. Fsw times are indicated. b) Cumulative distribution of MFS. Sets were classified into one of three categories: high stability (MFS>200h, light green); medium stability (20h<MFS<100h, green); low stability (MFS<2h, dark green). c) Simulations of three individual cells in one randomly selected parameter set of each category are shown starting from alternative initial conditions (1 and 2) indicated by arrows. Top: level of polymerases on each gene in three individual cells over time, bottom: Ratio of A and B polymerases along the gene in simulated cell 3 averaged over 1 h window.

5.2 Collisions and promoter repression are required for transcriptional memory

We next set out to simplify the model structure in order to identify the model reactions that are critical for transcriptional memory and to analyze their parameter values. Specifically, we wanted to understand which mechanisms of transcriptional interference are essential in order to generate transcriptional memory, and whether the description of transcription initiation and promoter repression could be further simplified. To this end, we developed five reduced model versions each either lacking one of the TI mechanisms, or simplifying transcription initiation or promoter re-activation into single-step reactions. They are summarized in Table 5.2. Each of the

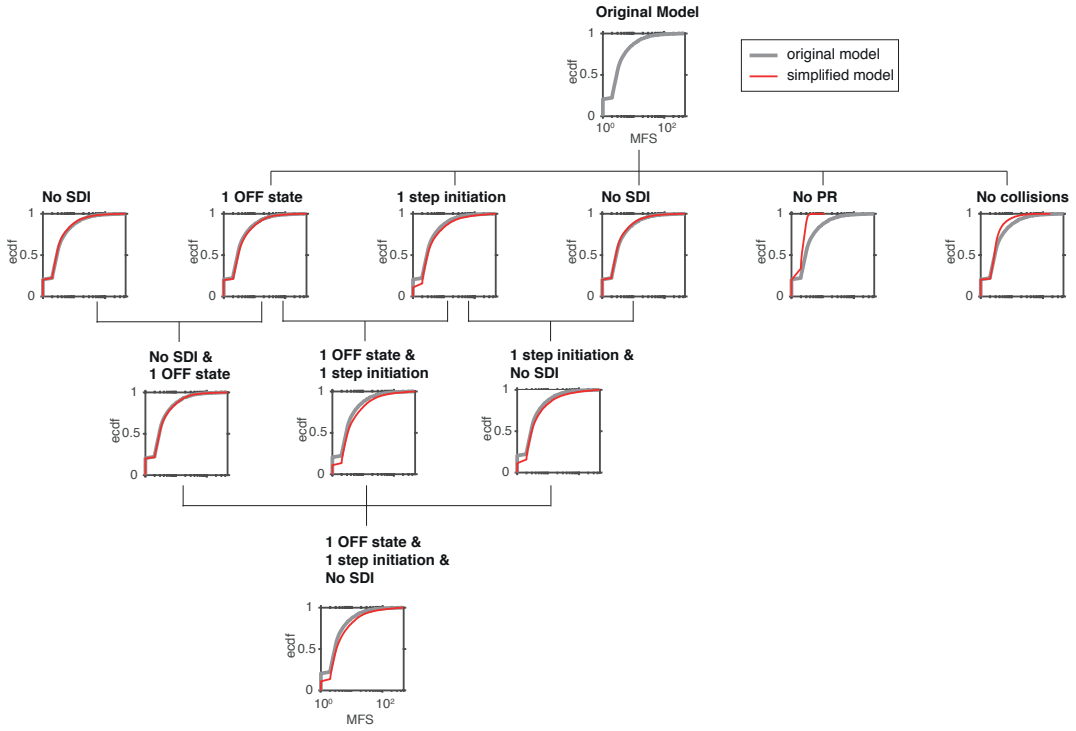

FIGURE 5.3: Cumulative distributions of minimal first-switching-time (MFS) as a measure for transcriptional stability in the original model (grey) and the reduced models (red). All simplifications that did not reduce the model's capacity to generate transcriptional memory were combined stepwise until maximally simplified model was obtained.

TABLE 5.2: Reduced antisense models. For each reduced model, the number of variable parameters, the mean MFS and the percentage of sets with MFS>200 h is shown.

Model	**# parameters**	**Mean MFS**	**% MFS>200h**
Original	10	7.4	0.43
No SDI	10	6.98	0.37
1 OFF state	9	7.67	0.33
1 step initiation	6	11.31	0.99
No PR	7	2.15	0
No collisions	10	3.80	0
No SDI & 1 OFF state	9	7.63	0.43
1 OFF state & 1 step initiation	5	11.86	1.09
1 step initiation & No SDI	6	11.95	1.14
1 OFF state & 1 step initiation & No SDI	5	11.74	0.95

reduced models was again simulated with 15,000 randomly sampled parameter sets and the minimal first-switching-times were compared to the original model (Fig 5.3, Table 5.2). We found that the removal of promoter repression (PR) and collisions (PC) strongly decreased the model's capability of maintaining alternative transcription states. However, SDI could be removed without a loss in maintenance potential. Similarily, transcription initiation could be simplified into a 1-step reaction, removing promoter-proximal pausing and termination. The number of promoter OFF states, interestingly, also had no effect on the maintenance potential of the model (Fig 5.3). To further simplify the model, we now tested all possible combinations of two simplifications that on their own did not decrease the model's capacity for transcriptional memory. We again found that none of these reduced model versions was restricted in its ability to maintain alternative

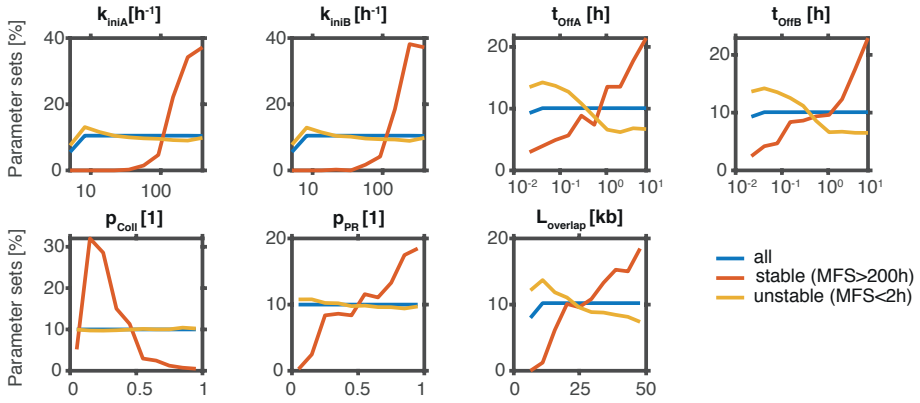

FIGURE 5.4: Distributions of parameter values in the simplified model across all tested parameter sets (blue), across all sets that could maintain the alternative transcription states for >200 h (stable, red) and <2 h (unstable, yellow). k_{iniA} and k_{iniB}: Initiation rates of A and B; t_{OffA} and t_{OffB}: Lifetime of pA_{OFF} and pB_{OFF}; p_{coll}: Probability for collision; p_{PR}: Probability for promoter repression; $L_{Overlap}$: Overlap length.

transcription states (Fig 5.3). Finally, we combined all three simplifications and found that even this maximally simplified model had not lost maintenance potential (Fig 5.3).

5.3 Parameter rules for memory

To understand how transcriptional memory can arise, we analyzed the parameter values that allow stable maintenance. We resimulated the maximally simplified model, with two additional parameters p_{Coll} and p_{PR} that modify the frequency of polymerase collisions and promoter repression, respectively, since it is difficult to estimate how often RNAP collisions and antisense RNAP induced promoter repression occur in living cells. Upon encounter of sense- and antisense RNAP, a collision event now occurs with probability p_{Coll}, and passing RNAPs induce repression of their antisense promoter with probability p_{PR}.

When analyzing the stable parameter sets we observed the following trends (Fig 5.4):

- The initiation rates k_{iniA} and k_{iniB} had to be high (>100 h^{-1}), indicating that strong promoters are required.
- Promoter repression must be stable, manifested in long half-lives of the repressed state t_{OffA} and t_{OffB} (>1 h).
- The distance between the two promoters must be long ($L_{Overlap} > 25$ kb).
- The probability for promoter repression must be high ($p_{PR} > 0.5$).
- The probability for collisions must be intermediate ($0.1 < p_{coll} < 0.5$).

To further understand these parameter requirements, we analyzed the model's sensitivity towards variation of each of the parameters. To this end, we randomly selected 100 stable sets

(MFS>200h) and systematically varied one parameter at a time, monitoring how a change in this parameter affected the transcriptional stability.

The initiation rates k_{iniA} and k_{iniB}

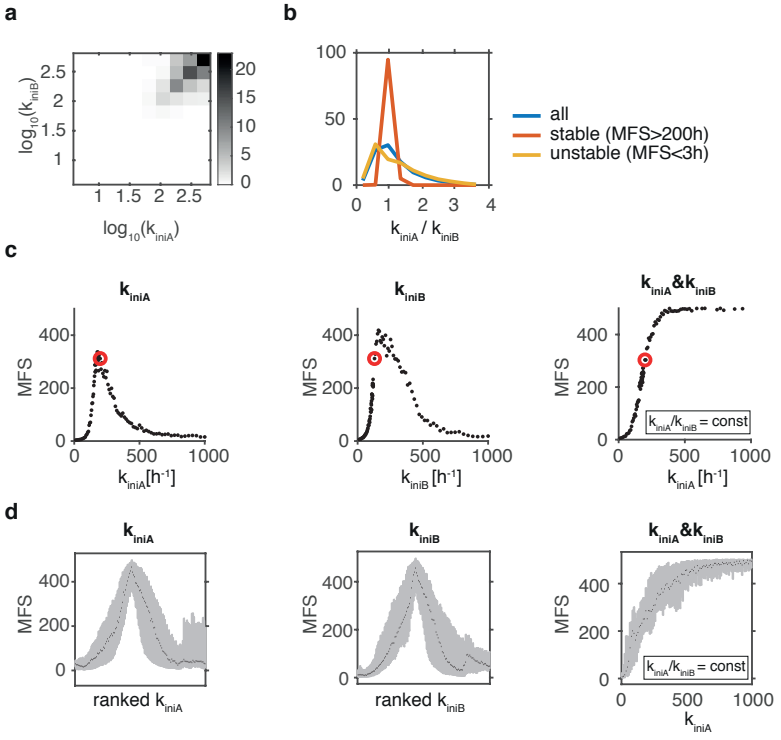

FIGURE 5.5: Effect of k_{ini} variation on the minimal first-switching-time (MFS). a) Heatmap showing the percentage of parameter sets that allow stable maintenance of alternative transcription states (MFS>200h) for different values of k_{iniA} and k_{iniB}. b) Distribution of k_{iniA}/k_{iniB} across all (blue), unstable (yellow) and stable (red) parameter sets. c) k_{iniA} (left), k_{iniB} (middle), or k_{iniA} and k_{iniB} (right) were varied in a randomly selected stable set (red circle) while all other parameters remained constant. The minimal first-switching-time is shown in dependence of k_{iniA} or k_{iniB}. d) k_{iniA} (left), k_{iniB} (middle), or k_{iniA} and k_{iniB} (right) were systematically varied in 100 randomly selected stable sets. Left and middle: The variations were ranked according to the value of the varied parameter. Right: Variations were ordered by the absolute value of k_{iniA}. Grey boxes display the 25th and 75th percentile and the black dot the median of the MFS in each bin.

Previous models of transcriptional interference have demonstrated that the strength of interference by collisions increases with increasing strength of the promoters [223]. Concordantly, we found that the initiation rates tended to be high in the stable sets, but we also found that they were positively correlated with each other and had a ratio close to one (Fig 5.5a,b). This suggests that absolute strength of transcription must be high and that promoter strength of the two genes must be balanced, likely because unbalanced initiation rates will allow the more aggressive promoter to quickly gain dominance over the weaker promoter independent of the initial conditions. Consistent with this, transcriptional stability is sensitive to small changes in either of the initiation rates k_{iniA} and k_{iniB}, as visualized for one example parameter set in Fig 5.5c. The same trend is observed across all tested sets if they are aligned relative to the k_{iniA} or k_{iniB} values for which they display maximal stability (Fig 5.5d). To validate that not only the ratio but also the absolute values of the initiation rates are crucial, we simultaneously varied both initiation rates keeping their ratio constant. As expected, transcriptional stability also decreases

if the initiation rates remain balanced but are decreased (Fig 5.5c,d, right). High initiation rates possibly help to filter out fluctuations that can lead to noise-induced switching of the locus.

Distance between promoters $L_{Overlap}$

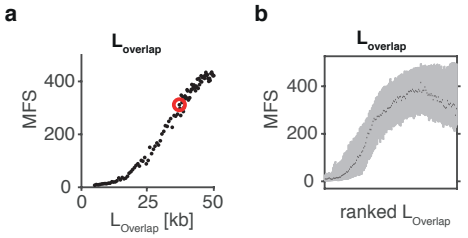

FIGURE 5.6: Effect of variation of the overlap length $L_{Overlap}$ on the MFS. a) $L_{Overlap}$ was varied in a randomly selected stable set (red circle) while all other parameters remained constant. The MFS is shown in dependence of $L_{Overlap}$. b) $L_{Overlap}$ was systematically varied in 100 randomly selected stable sets. The variations were ranked according to the value of the varied parameter. Grey boxes display the 25th and 75th percentile and the black dot the median of the MFS in each bin.

Systematic variation of the inter-promoter distance $L_{Overlap}$ in the 100 randomly selected stable sets revealed that transcriptional stability increased with increasing inter-promoter distance (Fig 5.6). A longer overlap likely aids the initially dominant gene to maintain its superiority, as it facilitates the removal of occasionally initiating antisense RNAPs by collisions by increasing the probability that two opposing RNAPs meet on the gene body [203]. Consistently, previous mathematical modeling has demonstrated that the interference by collisions felt by a gene increases with increasing inter-promoter distance [223].

Lifetime of the promoter OFF states t_{OffA} and t_{OffB}

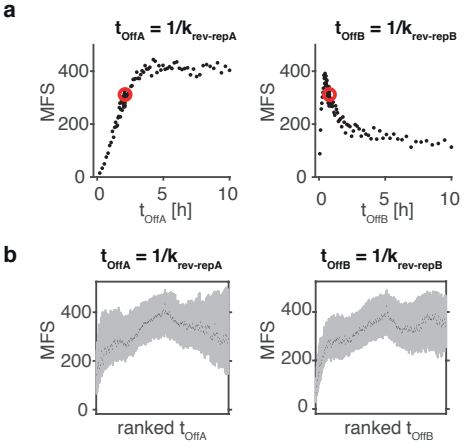

FIGURE 5.7: Effect of variation of the promoter OFF-state lifetimes on the MFS. a) t_{OffA} or t_{OffB} were varied in a randomly selected stable set (red circle) while all other parameters remained constant. The MFS is shown in dependence of t_{OffA} and t_{OffB}. b) t_{OffA} and t_{OffB} were systematically varied in 100 randomly selected stable sets. The variations were ranked according to the value of the varied parameter. Grey boxes display the 25th and 75th percentile and the black dot the median of the MFS in each bin.

Antisense RNAPs that transcribe over the sense promoter induce its transition to the OFF state.

Varying the lifetimes of the promoter OFF state had variable effects in different parameter sets. Consistently across all sets a too short life-time of the OFF state resulted in a dramatic loss of transcriptional stability (Fig 5.7b), in agreement with our previous observation that promoter repression was strictly required to generate longer transcriptional memory. Increasing the life-time had variable effects but overall higher stability of the repressed state resulted in longer transcriptional memory (Fig 5.7b). In some sets (e.g. Fig 5.7a) modifying t_{OffA} or t_{OffB} had asymmetric effects, such that transcriptional stability increased if t_{OffA} was increased, but decreased upon t_{OffB} increase (or vice versa). One possible reason for this could be that the total interference of collisions and PR together, exerted by each gene on its convergent promoter was not completely balanced in the original parameter set. Tilting the relative strength of the two genes to a more balanced value (e.g. by increasing t_{OffA} or decreasing t_{OffB} (Fig 5.7a)) might therefore still increase transcriptional stability.

Probability for promoter repression p_{PR}

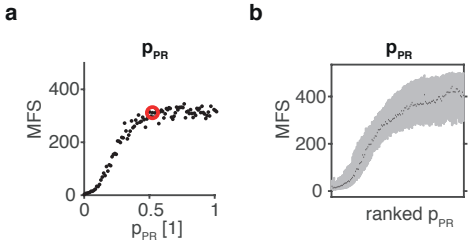

FIGURE 5.8: Effect of variation of the probability for promoter repression p_{PR} on the MFS. a) p_{PR} was varied in a randomly selected stable set (red circle) while all other parameters remained constant. The MFS is shown in dependence of p_{PR}. b) p_{PR} was systematically varied in 100 randomly selected stable sets. The variations were ranked according to the value of the varied parameter. Grey boxes display the 25th and 75th percentile and the black dot the median of the MFS in each bin.

Transcriptional stability increases with increasing probability for promoter repression by antisense RNAPs (Fig 5.8). Frequent promoter repression acts in favor of the dominant gene reinforcing repression of the OFF gene. If both promoters are in the ON state interference is governed by (symmetrically acting) collisions until one gene manages to reach and repress its antisense promoter at which point the balance is tilted in favor of this gene. It can maintain its superiority by preventing initiation of any new antisense RNAPs. Thus a higher probability for promoter repression helps the currently dominant gene to maintain its superiority.

Probability for collisions p_{Coll}

In vivo, it is unclear how frequently collisions between antisense RNAPs occur, and whether mechanisms exist that facilitate bypassing [154]. We therefore also investigated whether the probability with which a collision event occurs upon encounter of sense and antisense RNAPs (p_{Coll}) affects transcriptional stability. Curiously, we had observed that the most stable sets tended to have intermediate p_{Coll} values (see Fig 5.4). Systematic variation of p_{Coll} confirmed that there is an optimal p_{Coll} value in all tested parameter sets (Fig 5.9a,b).

We hypothesized that this optimum might exist because the initiation rates of the two convergent promoters are not completely balanced. As a consequence, higher p_{Coll} values will favor

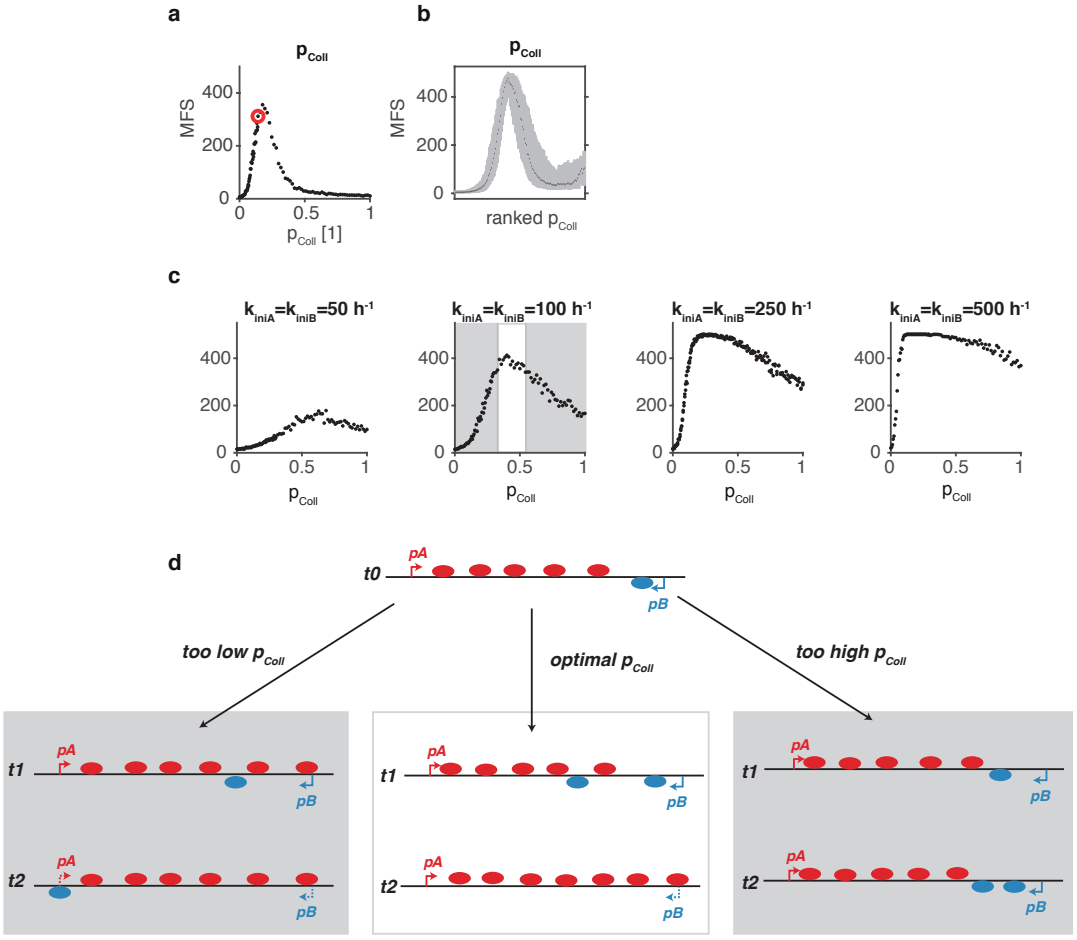

FIGURE 5.9: Effect of variation of the probability for polymerase collisions p_{Coll} on the MFS. a) p_{Coll} was varied in a randomly selected stable set (red circle) while all other parameters remained constant. The MFS is shown in dependence of p_{Coll}. b) p_{Coll} was systematically varied in 100 randomly selected stable sets. The variations were ranked according to the value of the varied parameter. Grey boxes display the 25th and 75th percentile and the black dot the median of the MFS in each bin. c) Systematic variation of p_{Coll} in a symmetric parameter set ($k_{iniA} = k_{iniB}$; $t_{OffA} = t_{OffB} = 5$h; $p_{PR} = 1$; $L_{Overlap} = 25$ kb) for different values of the initiation rates k_{iniA} and k_{iniB}. d) Scheme illustrating a possible reason for the p_{Coll} optimum. Intermediate p_{Coll} values might be optimal to cope with occasionally initiating antisense poymerases. If p_{Coll} is too low, initiated antisense polymerases can reach and repress the sense promoter escaping their removal upon collision (left). If p_{Coll} is too high, too many sense polymerases are removed due to collisions, prolonging the time until re-repression of the antisense promoter and thereby allowing the production of more antisense polymerases (right). When p_{Coll} is optimal, sense polymerases re-repress the antisense promoter reasonably fast, and at the same time manage to reliably remove any initiated antisense polymerases before they can reach and repress the sense promoter (middle).

the gene with the stronger promoter. To test this hypothesis, we analyzed the effect of p_{Coll} variation in a few completely symmetric parameter sets (k_{iniA}=k_{iniB} and t_{OffA}=t_{OffB}). Surprisingly, also in these symmetric sets an optimal p_{Coll} value existed (Fig 5.9c). When inspecting a few example trajectories, we saw that usually the transcription state is retained until the initially inferior gene manages to repress the promoter of the initially dominant gene, and this is prevented most efficiently at intermediate p_{Coll} values. Why is that? The higher p_{Coll}, the more antisense RNAPs are removed from the gene, but also the more of a gene's own RNAPs are lost to collisions. Figure 5.9d schematically illustrates how intermediate p_{Coll} values might allow for the optimal trade-off between minimizing self-repression and maximizing repression of the antisense gene. If p_{Coll} is too low, initiated antisense RNAPs can easily reach and repress the sense promoter escaping removal by collisions (Fig 5.9d, left). If p_{Coll} is too high, too many of the genes own RNAPs are lost to collisions. This prolongs the time until re-repression of the

antisense promoter and thereby allows the production of more antisense RNAPs (Fig 5.9d,right). When p_{Coll} is optimal, sense polymerases re-repress the antisense promoter reasonably fast, and at the same time manage to reliably remove any initiated antisense polymerases before they can reach and repress the sense promoter (Fig 5.9d, middle).

In summary, we found that transcriptional interference can maintain alternative transcription states at an antisense locus for several days given that the locus fulfills some requirements with respect to its architecture and the kinetic properties of the promoters. Our analysis predicts that stability of the transcription state requires frequent and stable promoter repression by antisense transcription, occasional RNAP collisions, strong balanced promoters and a long inter-promoter distance. The frequency of collisions and promoter repression by antisense RNAPs are likely inherent characteristics of mammalian transcription and therefore difficult to perturb experimentally. Similarly, the stability of promoter repression cannot be modulated in a straightforward manner. Promoter strength, and inter-promoter distance, on the other hand, can be modulated in an experimental system to test central predictions of our antisense model (see section 6.6).

6 Discussion

As the characterization of the nascent transcriptome advances it is becoming increasingly clear that antisense transcription is ubiquitously present across genomes of bacteria, archaea and eukaryotes. But apart from a few well-characterized examples we are still in the dark about its functions. Here, we have investigated the potential of antisense loci to stably retain the transcription state they have acquired due to a signal encountered in their past.

6.1 Making stable memories

Fundamentally, a memory of an expression state can either arise through an irreversible change, or through active maintenance of alternative expression states. Transcription is regulated by biochemical reactions that are reversible, except for changes to the DNA sequence, which are faithfully inherited throughout cell divisions. A classical paradigm for the storage of acquired information through irreversible changes to the DNA sequence is the adaptive immune system. Here, the enourmous diversity in the antigen receptor binding region between different lymphocytes is generated through random combinatorial rearragements of V, (D) and J segments that determine the variable region of the antigen receptor [268]. The transient action of sequence-specific endonucleases initiates a recombination reaction, that results in a permanent change to the DNA sequence. Thus, a particular lymphocyte stores the memory of its antigen receptor as a permanent change in its DNA sequence. The irreversibility of changes to the DNA sequence is also exploited by molecular biologists to create expression memory. The Cre-LoxP system for instance allows researchers to permanently switch expression of a gene ON or OFF in a subpopulation of cells, via a transient (cell type-specific) signal that temporarily activates Cre recombinase. Cre then permanently excises the DNA between the two *LoxP* sites, resulting in permanent activation or inactivation of the target gene depending on the experimental design.

If not locked in by an irreversible change, active self-reinforcing mechanisms must be in place to maintain expression memory. Positive feedbacks can indeed actively maintain memory of a transient stimulus, as has been demonstrated in engineered bistable circuits in *S. cerevisiae* and *E. coli* [122, 161]. However these synthetic approaches have also shown that stochastic effects can cause frequent switches between the states [161]. To generate stable memory, robustness against such fluctuations is required. The bistable regime should extend over a broad enough range of cellular states to buffer these random fluctuations. In biology, this is often achieved through the coupled action of multiple feedbacks [160, 163, 269, 270].

Short-term memory of an expression state might not even require bistability: Systems that display ultrasensitive responses might also appear to demonstrate short-term memory, even if they are not actually bistable, if the kinetics with which the system reaches its new steady state are slow. Experimentally, it might be difficult to distinguish whether a system is truely bistable or simply slow to react.

Since here we are interested in the biological phenomenon, we have explored the network's behaviour empirically taking into account stochastic fluctuations. To this end, we have scanned a range of biologically relevant parameters and probed for the ability to retain alternative transcription states by exposing the network to different initial conditions. We found that transcriptional memory over several days can arise at antisense loci, given that the locus fulfills some key requirements regarding genomic architecture and transcriptional properties, which will be discussed in the next section.

6.2 Characteristics of mammalian transcription

In our model, lasting memory is produced by antisense pairs that have:

I A locus architecture with a long inter-promoter distance ($>$ 20 kb) and with both promoters residing inside the overlap, to allow mutual regulation of the promoter-proximal chromatin environment.

II Potent promoters of comparable strength.

III Stable mutual repression of transcription initiation through the establishment of a repressed chromatin environment at the antisense promoter.

Are these requirements compatible with the characteristics of mammalian transcription?

I) Architecture of mammalian antisense pairs
Antisense transcription has been estimated to occur at ∼30% to 70% of annotated transcription units in mammalian genomes [196, 271–273]. Antisense transcription frequently arises within the known genes and extends through their promoter, so that both transcription start sites (TSS) lie within the overlap [273–275]. In HeLa cells, the median distance between the promoters of simultaneuosly expressed sense and antisense transcription units has been estimated to lie somewhere in the range of a few hundred bp, and is thus much shorter than the overlaps that we find to be required for stable transcriptional memory [273]. However, at individual alleles antisense pairs that mediate transcriptional switches are not expected to be transcribed simultaneously. Therefore, profiling of transcription across different conditions (e.g. different tissues) is needed to systematically identify genomic locations that are transcribed from opposite strands in different cellular states. In the meantime, it can be informative to look at genome-wide length distribution of all annotated overlapping transcription units. In the mouse genome, the mean overlap between two genes encoded on opposite strands is almost 8 kb, and there are roughly 1000 annotated convergent transcription units that overlap for more than 10 kb[1]. This, of course,

[1] analyzed based on the mm10 annotation from UCSC [276]

does not mean that the two convergent promoters are ever simultaneously active in any cellular state, but it still gives an indication about the general genomic organization. In principle, the mammalian genome does thus harbour a number of antisense pairs that are long enough to generate transcriptional memory by mutual repression.

It is also very well possible that we still underestimate the number of convergent transcription units: Most of the existing genome-wide datasets that have been used to analyze antisense transcription were acquired with techniques that are strongly biased towards the detection of highly expressed, stable transcripts, such as conventional RNA-Seq or microarrays. Since many antisense genes encode cryptic unstable lncRNAs, they are likely to escape detection by these techniques [277]. Especially in cases where the antisense gene exerts its function through the act of transcription rather than the RNA itself, as has been investigated here, the resulting transcripts might be unstable. Techniques that profile nascent transcription instead of transcript levels are thus better suited for the analysis of antisense transcription, for instance through metabolic labeling of newly synthesized RNAs (TT-Seq [278]), or by measurements of PolII density (NET-Seq [275, 279, 280]). A study that has globally profiled nascent transcription in yeast indeed finds three trends that are compatible with our model [277]:
- Sense and antisense transcription are globally anticorrelated.
- The level of sense transcription decreases as the overlap with the antisense transcription unit increases.
- Genes with antisense transcription extending into their TSS tend to be transcribed at lower levels than gene without antisense transcription covering their TSS.

As our knowledge of the nascent transcriptome increases, and as more data that compare transcription across different conditions are acquired, we might identify more genomic locations with transcription patterns that could indicate the presence of transcriptional switches or memory mediated by antisense transcription. Candidate genomic locations should fulfill the following criteria:
- Both strands should be capable of initiating transcription.
- Sense and antisense transcription should be anticorrelated on single alleles.
- Across different conditions the strand that dominates transcription should switch.

II) High and balanced initiation frequency of sense and antisense promoter
In our model analysis, memory in the range of several days arises when sense and antisense promoters have initiation frequencies above 100 RNAPs per hour. Experimental estimates of global initiation frequencies are scarce but a seminal study by the Cramer lab has used measurements of nascent transcription to estimate absolute global initiation frequencies in human cells [281]. They measure a mean initiation frequency of 160 RNAPs/h across all detected genes, which is in the range of the initiation frequencies that we find to be able to generate transcriptional memory that lasts for several days.

III) Ability for mutual promoter repression
Global analysis of the effect of antisense transcription on chromatin architecture in yeast has shown that antisense transcription causes changes to the sense promoter-proximal chromatin. It increases the nucleosome occupancy and affects histone modifications around the sense TSS

[273, 282, 283]. This antisense transcription-induced chromatin signature seems to be conserved between yeast and human, and indeed seems to reduce initiation of transcription at the sense promoter [273, 282, 284]. To dissect exactly how antisense transcription modulates sense transcription dynamics, i.e. initiation, elongation and degradation rates, measurements that profile transcript levels, RNAP occupancy and nascent transcription must be combined.

Misregulated antisense transcription that results in stable repression of the sense gene is also implicated in disease, such as the ANRIL antisense RNA that stably silences a tumor suppressor locus through the recruitment of polycomb [242, 285], or the HBA2 promoter that is silenced through the recruitment of CpG methylation by an erroneously acquired antisense transcript in a genetic form of anemia [237]. In addition, many imprinted gene clusters are also silenced in a parental-specific manner by antisense transcription that recruits histone modifying enzymes or DNA methyltransferases which stabilize the imprint and guarantee its faithfull propagation throughout cell divisions [130, 239, 246].

Our model predicts that antisense pairs which do not perfectly fulfill all requirements (i.e. shorter overlap, weaker promoters, or less stable mutual promoter repression) can also memorize transcription states, albeit for shorter time scales.

These considerations argue that in principle antisense loci that are capable of mediating transcriptional memory could exist in mammalian genomes, but that we do need appropriate techniques and experimental designs to identify such antisense pairs. Specifically, in addition to transcript levels, nascent transcription should be measured across different conditions to identify such antisense pairs that switch their expression state between conditions.

6.3 Non-linear responses and memory by antisense transcription

Antisense transcription has already been invoked in producing non-linear responses in other biological contexts, some of which will be discussed in this section. These examples demonstrate that TI between antisense transcripts can filter out low level sense transcription and generate switch-like expression patterns. Usually, one of the two convergent promoters dominates transcription in the default state, and the locus can be switched by binding of a transcription factor that activates the other strand, yet without inducing memory: In the absence of the transcription factor the default strand becomes dominant again. One such example is entry into meiosis in yeast which requires the transcription of the initiator of meiosis genes (IME). One of them, IME4, has an antisense transcript, and the two genes repress each other by transcriptional interference. The antisense transcript is the "default" transcript, being driven by the stronger promoter and thereby dominating transcription at the locus in haploids. To switch the locus, constitutive antisense transcription must be repressed, which can happen in a/α diploids by binding of the a1/α2 heterodimer which represses antisense transcription and thereby allows full-length transcription from the weaker sense promoter [286]. In this way, antisense transcription reliably suppresses noisy sense transcription that could otherwise cause inappropriate entry into meiosis in haploids.

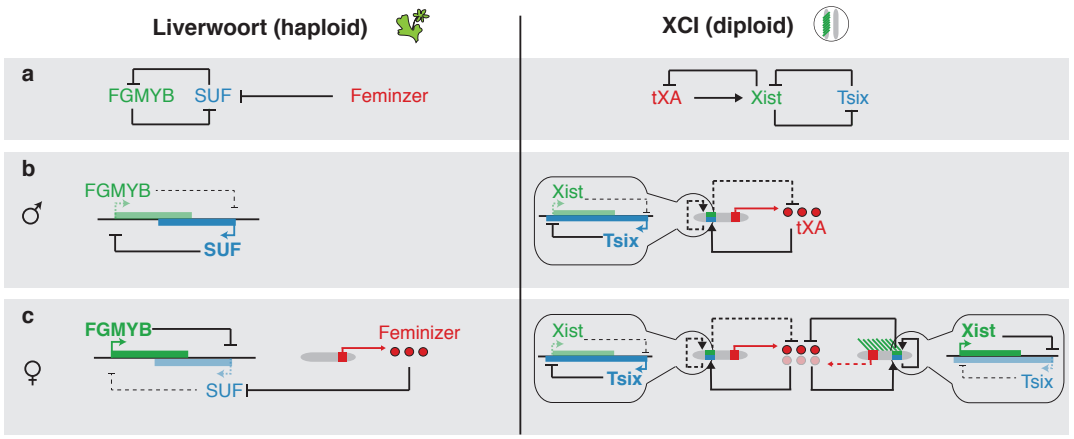

FIGURE 6.1: A single autosomal antisense locus controls sexual dimorphism in the liverwort by acting as a switch that is controlled through an X-encoded regulator (left), reminiscent of our XCI model (right). a) Network modules governing the two regulations. In both cases, antisense repression generates switch-like activation of the sense strand (green) triggered by the expression of an X-encoded regulator (red). Coordination of the Xist expression pattern on the two alleles in diploid mammals requires silencing of tXA, and necessitates local memory of Xist expression on the Xi mediated by antisense repression. b-c) Expression pattern of the antisense locus in males (b) and females (c). In the haploid liverwort the X-encoded regulator (red circles) is expressed in an all-or-nothing manner between the sexes. In males, MpFGMYB transcription is suppressed by the constitutively expressed antisense gene SUF (b). An X-encoded 'feminzer' factor suppresses SUF transcription in females, allowing MpFGMYB expression, which then activates downstream genes that promote female differentiation [287] (c). In mammals tXA levels vary two-fold between males and females, and antisense repression locks in the Xist expression state on Xa and Xi upon tXA silencing in females.

Transcriptional interference thus creates an all-or-nothing switch that is controlled by a *trans*-acting factor.

Another remarkable example of an all-or-nothing expression pattern mediated by antisense transcription is the control of sexual dimorphism in the liverwort, where a single autosomal antisense locus determines whether the plant produces male or female reproductive organs [287]. This example is especially worth discussing because of its parallels to XCI regulation: In both cases the transcription state of the antisense locus is switched in a sex-specific manner by an X chromosome-encoded regulator. In regulation of XCI this regulator is the tXA factor, in regulation of liverwoort sexual dimorphism it is the *feminizer* gene (Fig 6.1a). In the liverwort, one strand of the antisense locus encodes a MYB-domain transcription factor (MpFGMYB), which is required for the generation of female reproductive organs, on the other strand the locus encodes a *cis*-acting repressor of the sense transcription (the lncRNA SUF). The two genes are expressed in a mutually exclusive and sex-specific manner, with MpFGMYB being exclusively expressed in female plants, and SUF being exclusively expressed in male plants, reminiscent of the expression pattern of Xist and Tsix on Xi and Xa (Fig 6.1 b-c). The current model postulates that SUF antisense transcription is suppressed specifically in females by the X chromosome-encoded (and thus female-specifically) expressed *feminizer* gene (Fig 6.1c), reminiscent to tXA-mediated control of Xist switch-on. In both regulations, this *trans*-acting factor switches the expression state of the antisense locus thereby controlling the downstream regulation (Fig 6.1a). In contrast to the regulation of XCI, though, the *MpFGMYB/SUF* antisense locus does not need to convert quantitative into qualitative information as the organism is haploid and expression of the *feminizer* gene is thus binary (expressed in females, not expressed in males), rather than graded as

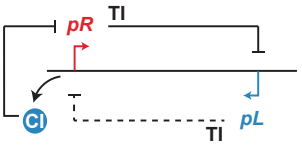

FIGURE 6.2: Convergently oriented lytic (pR) and lysogenic (pL) promoters control the lysis/lysogeny decision in bacteriophage 186 through mutual inhibition on RNAP and protein level.

is the expression of tXA in mammals (2x tXA in females, 1x tXA in males) (Fig 6.1b-c, compare left to right). Further, there must be no coordination of the locus' expression state on different alleles. Therefore, no negative feedback reducing the expression of the X-encoded activator is required. As a consequence, the locus does not need to acquire a memory of its transcription state, as the *feminizer* gene can potentially be continuously expressed to activate MpFGMYB transcription.

During XCI, however, each X chromosome must acquire a memory of its transcription state, which, as the first part of this thesis has shown, could be mediated by Xist/Tsix antisense transcription. Stable maintenance of alternative Xist transcription states on Xa and Xi can arise if the *Tsix* promoter is silenced by Xist RNA, and Tsix RNAPs are capable of repressing the *Xist* promoter. This constitutes a double negative feedback whose two arms are mediated on different levels: Tsix dominates repression on the level of transcription and Xist dominates repression on the posttranscriptional level. Tsix possesses the stronger promoter, allowing it to dominate transcription on the Xa where it constantly reinforces repression of the *Xist* promoter. On the Xi, the Xist RNA mediates silencing of the *Tsix* promoter. Periods of lower transcriptional activity at the *Xist* promoter can be overcome, as the Xist RNA has a half-life of several hours and can therefore buffer noise in transcription. Thus, mutual repression between Xist and Tsix is exerted on different levels: Tsix dominates RNAP-mediated repression, and Xist dominates RNA-mediated repression.

Such asymmetric multilayer repression is reminiscent of another regulation between a pair of antisense genes that displays bistability: The lysis/lysogeny switch in bacteriophage 186, where the lytic and lysogenic promoters are oriented convergently such that antisense transcription can contribute to their mutual repression. One of the two genes has the stronger promoter and dominates the fast-acting TI-mediated repression, while the weaker gene encodes a protein that can repress the stronger promoter, and thus dominates slow protein-mediated repression (Fig 6.2). However, it is probably rather the exception than the rule that the gene products encoded by one strand of an antisense locus are capable of regulating their convergent promoter. Our general antisense model demonstrates that transcriptional memory could also arise at convergent transcription units that solely interfere with each other's transcription by RNAP-mediated mechanisms. In our model, the initiation rates do not only determine the chances of a gene to survive the collisions but they also determine the frequency with which each gene can reinforce repression of its antisense promoter. As a consequence, the strength of the two promoters must be comparable, to balance the two legs of the double negative feedback. This contrasts with the *Xist/Tsix* locus where *Xist* relies on stable RNA-mediated repression of *Tsix* on the Xi, while *Tsix* has the stronger promoter, which allows it to dominate collisions and repress the *Xist* promoter on the Xa.

These examples demonstrate that antisense transcription can ensure switch-like activation of

the sense transcript. In some cases, antisense transcription also contributes to the establishment of expression memory. These examples usually involve regulation on multiple layers, where the RNA transcript or the protein encoded by one of the strands regulates expression of the other strand. In our model we have shown that epigenetic memory could in principle also arise by transcriptional interference alone, without feedbacks involving the encoded RNAs or proteins. Thus, transcriptional memory can arise, when both genes are able to repress each other with comparable strength. If mutual repression between the two genes is exerted on multiple layers, memory can also arise if one gene dominates repression on one layer, and the other dominates repression on another layer. In the next section, I will sketch a few regulatory scenarios in which such transcriptional memory could be advantageous.

6.4 Potential functions of antisense transcription-mediated memory

There are many situations in biology where a transient stimulus must be converted into a sustained response. These range from short- to long-term memories. A classical example for long-term memory is cell differentiation into distinct cell types, where a cell must remember its identity, defined by its transcriptional program, throughout its complete lifetime. Such stable cell fate decisions are often governed by mutual inhibition between two opposing fate-determining transcription factors that repress each other by various molecular mechanisms, often involving protein-protein interactions [124]. A shorter-term memory can be observed in the immune system, for instance during antigenic stimulation of T cells: CD4+ Th2 lymphocytes that are stimulated by an antigen transiently activate IL-4 expression, to promote an antibody-mediated immune reaction [288]. The lymphocytes that have activated IL-4 upon the initial stimulation, retain a short-term memory in the range of days, and are more likely to reactivate IL-4 upon a second stimulation [289]. In HeLa cells, approximately half of the genes that respond to IFN gamma are induced faster or stronger if they have previously been exposed to IFN gamma stimulation [290, 291]. Antisense-mediated transcriptional memory could be involved in mediating such shorter-term memories, that last for hours to days. Below, three possible scenarios, in which antisense transcription-mediated memory could be of use, are detailed.

Noise filtering

Cells see variable environments and must be able to distinguish stochastic fluctuations from true signals. Short-term memory of the transcription state can provide a useful buffer against fluctuations in the signals that a cell receives by filtering out signal changes below a specific magnitude or duration. This can be of use especially for cells that are confronted with a quickly changing environment, i.e. all cells that do not form a solid tissue but are in constant movement (e.g. blood cells). In this way, periods of temporarily low or absent activation signals can be overcome, or conversely, spurious activation signals can be filtered out.

Signal extension

Possibly, the most intuitive function of transcriptional memory is the production of a prolonged transcriptional response to a transient stimulus. A transient stimulus might activate the antisense strand and thereby repress sense transcription. The locus then retains its switched transcription

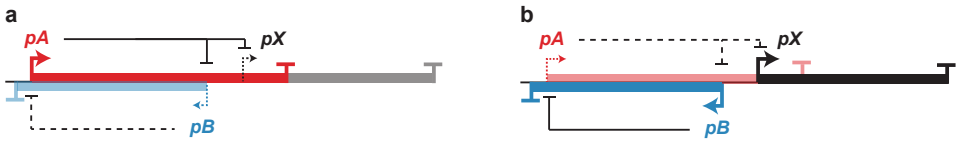

FIGURE 6.3: Antisense toggle switch regulating the expression of a neighboring gene X. Gene A overlaps with pX and represses gene X by transcriptional interference. The locus can exist in two alternative transcription states: a) Gene A is active and represses its antisense gene B and the target gene X. b) Gene B is active and represses gene A, allowing gene X to be expressed. If the antisense locus retains a memory of its transcription state, this memory will be transmitted to the target gene.

state after the initial stimulus is gone. In cases that require very stable transcriptional memory, antisense transcription might ensure initial silencing of the sense gene, which could later be stabilized through additional mechansims such as the recruitment of repressive chromatin modifications. In plants silencing of the FLOWERING LOCUS C (FLC) is initiated by upregulation of its antisense transcript COOLAIR, and is then stabilized by the recruitment of polycomb which establishes H3K27me3. Exposure to cold induces transcription of COOLAIR which represses FLC transiently, and therefore is essential in the early stages of silencing. Recruitment of polycomb then confers the epigentic memory that ensures that FLC remains stably silenced even after return to the warm, so that flowering can initiate [249, 250]. Similarily, Tsix transcription initially represses Xist on the active X, while other mechanisms such as DNA methylation take over in somatic cells where Tsix is no longer expressed [61, 62].

Signal memory

The term "transcriptional memory" is most commonly used to describe the priming of a gene for reactivation. Such transcriptional priming allows cells to *learn* from a previous transient stimulus and to respond in a more efficient manner when the same stimulus is encountered again [291, 292]. Such behaviour is well known from plants which use transcriptional priming to cope with the various environmental stresses they encounter [293]. It is conceivable that antisense transcription could encode such memory of a previous stimulation. An antisense locus could, for instance, act as an upstream regulator of a target gene that displays transcriptional memory of a previous activation. We could imagine an antisense locus encoded in close proximity to our target gene (Fig 6.3). One of the two convergent transcription units overlaps the target gene and represses it by transcriptional interference. Now, two transcriptional configurations are possible: One in which this strand is active and the target gene is repressed (Fig 6.3a), and a second one in which this strand is repressed by its convergent transcription unit, which allows activation of the target gene (Fig 6.3b).

Such regulatory examples have been documented in yeast, where a toggle switch between a pair of antisense genes upstream of the *FLO11* locus controls FLO11 expression in an all-or-nothing manner [294]. If the antisense locus has been switched to the favorable transcriptional configuration upon the initial stimulation, it will retain a memory of its transcription state for some time, and upon secondary stimulation allow for faster activation of the target gene.

Synthetic logic gates and memory devices

The transcriptional behavior of antisense loci could also provide a useful tool in synthetic biology to construct memory devices or logic operators. A synthetic memory circuit that can remember a past stimulus can have useful biotechnological applications, for instance for the sustained production of a target protein after induction by a transient stimulus.

Logic operations, on the other hand, are a part of gene regulatory networks and help cells to integrate many different environmental or cellular signals. In synthetic biology they can be useful for generating precise qualitative behavior in response to specific input signals. A number of synthetic gene circuits have been constructed already that contain digital logic operators [122, 295, 296]. These logic operators produce qualitatively different behavior in response to different combinations of signals. At an antisense locus, the presence of signal 1 might result in transcription of the sense strand, while the presence of signal 2 results in transcription of the antisense strand, a feature that can be exploited for instance to build NOT IF logic gates. Such logic operators are able to respond to different small molecule inputs and thereby enable logical transcription control in mammalian cells.

6.5 Limitations of the study

Intrigued by the possible role of Tsix in setting up and maintaining alternative Xist expression states we have set out to analyze the general potential of antisense transcription to maintain expression memory. Whether at all, and how frequently such regulatory scenarios where antisense transcription maintains a specific expression state or primes a locus for re-activation exist, remains an open question that needs to be addressed with appropriate techniques such as profiling of nascent transcription across various conditions: Antisense transcription involved in cellular switches would be expected to display transcription of different strands in different conditions. How relevant this mode of regulation really is in mammalian transcription regulation thus remains to be seen.

It will be interesting to analyze the effect of discontinous bursty transcription initiation on the stability of alternative transcription states in our antisense model. Many genes are transcribed in short intense bursts rather than in a continuous manner, where average waiting times between bursts can be in the range of several hours [151, 213, 216]. Such spontaneuos switches of the promoter to the inactive state could complicate the maintenance of alternative transcription states at antisense loci: If the promoter of the currently dominant gene spontaneuosly switches off and remains in the inactive state for some time, this will relieve repression on the antisense promoter giving it the opportunity to reactivate and repress the sense promoter. Likely, this would result in more frequent switches of the locus' transcription state. Burst kinetics however vary between genes, and can approximate continuous transcription [216, 297]. TATA-box containing promoters e.g. tend to be highly permissive to transcription with long transcriptionally active (ON) periods, and high RNAP firing rates [297]. For genes that are transcribed without long pauses in transcription initiation, mutual RNAP-mediated repression between convergent promoters remains a plausible mechanism for the generation of transcriptional memory. At antisense loci with longer inactive promoter periods, possibly only shorter term transcriptional memory can arise.

At the *Xist/Tsix* locus bursty transcription could also affect the stability of alternative Xist expression states. Xist in fact seems to be transcribed in bursts: In RNA FISH data of differentiating mESCs some alleles with an exonic Xist RNA signal do not have an intronic signal, indicating that Xist has been transcribed previously but is not being transcribed at the moment.

In addition, Xist RNA distributions across single cells are broader than expected from non-bursty transcription in a homogeneous cell population. How much of this variability stems from extrinsic factors (e.g. differences in cell size, cell cycle stage, pluripotency state, tXA expression level) would need to be dissected, for instance by quantifying the variability in expression between alleles of the same cell [298]. Spontaneous switches of the *Xist* promoter to an inactive state might complicate the establishment of the Xist^{high} expression state on the Xi. Xist RNA-mediated silencing of Tsix happens with a delay. During this time, Xist solely relies on TI to repress Tsix. Therefore periods of promoter inactivity make it vulnerable to Tsix reactivation and destabilize the Xist^{high} state. Depending on the duration of the inactive periods, Xist might require an additional (immediately acting) mechanism, to bridge the delay before complete Xist RNA-mediated silencing of *Tsix* sets in. Possibly, certain Tsix TSSes are quickly silenced by Xist RNA, reducing the strength of total Tsix transcription, while others take longer to be inactivated. This notion is consistent with data profiling nascent transcription, that indicate that Tsix transcription dynamically arises from different promoters over differentiation, and between XX and XO cells[2]. It will be interesting to quantitatively fit single cell RNA distributions of Xist with a model that accounts for transcriptional bursting.

6.6 Outlook - A synthetic antisense locus

We have developed and analyzed a theoretical framework that describes a circuitry of mutual repression between a pair of antisense genes. To now test the model, we set out to synthetically construct an antisense locus in mESCs, to experimentally isolate and thereby define key players.

In the experimental assay, we will probe for transcriptional memory in a manner analogous to our theoretical analysis, by subjecting the system to different initial conditions and analyzing whether it gets locked in different stable transcription states, and how this behaviour depends on the parameters that have been identified as critical for transcriptional memory in our theoretical analysis. To this end, we have established a platform that allows straighforward stable integration of alternative antisense constructs into a neutral chromosomal target locus in mESCs. The experimental design also incorporates the means to trigger transcription from sense- or antisense strand independently, allowing us to subject the system to different initial conditions and possibly even to flip the system between alternative transcription states (Fig 6.4). A more detailed description of the steps that have been taken towards the construction of this experimental system, including the means to trigger sense or antisense transcription, the integration system, and the means to monitor the locus' transcription state can be found in Appendix C.

Fig 6.4 summarizes the design of the experimental assay. An antisense locus that displays transcriptional memory is expected to behave in the following way: In the absence of inducers, two transcription states of the antisense locus are possible. One in which gene A is strongly transcribed and represses transcription of gene B, and one in which gene B is transcribed and represses transcription of gene A. The locus can be forced into a defined initial transcription state by transient chemical induction of either gene A or gene B. Upon removal of the trigger, the locus should stably maintain its transcription state. By then giving the opposing trigger, it

[2]Till Schwämmle, Rutger Gjaltema, Edda Schulz, personal communication

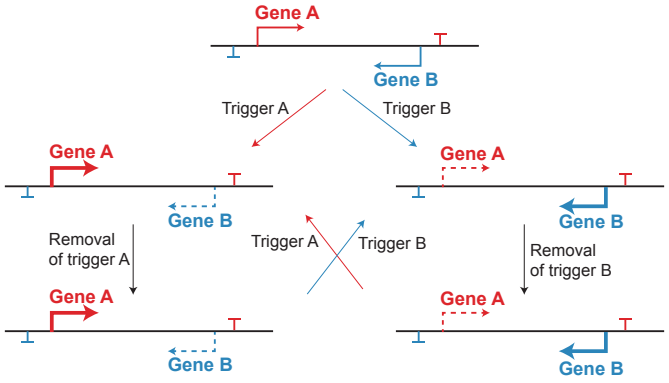

FIGURE 6.4: Experimental strategy to probe for transcriptional memory. Transcription from both strands of the antisense locus is separately inducible. This allows us to force the antisense locus into two alternative initial conditions, where either gene *A* or gene *B* dominate transcription. Upon removal of the trigger it can be analyzed for how long the locus remembers its transcription state. The state of the system can then possibly also be switched by giving the opposing trigger.

should be possible to switch the transcription state of the locus, which, again, should then be stably maintained even after removal of the inducer (Fig 6.4).

We aim to construct an antisense locus that indeed displays transcriptional memory, and to then perturb specific characteristics of the locus one at a time, evaluating their impact on transcriptional memory. Specifically, we plan to modify the following features that are predicted to be essential for transcriptional memory by the theoretical analysis: (i) strong and balanced promoters, (ii) a long inter-promoter distance, (iii) mutual promoter repression (Fig 6.5a). To understand the impact of promoter strength, i.e. initiation rates, we will compare our default construct where both sense and antisense strand are transcribed under a strong promoter (e.g. EF1a), to a construct, in which one of the two strands has a weaker promoter (e.g. Pgk), or a construct in which both strands have weaker promoters (Fig 6.5b). To evaluate the impact of the overlap length we will compare constructs with different length of intronic sequence integrated between the two promoters (Fig 6.5c). To test the importance of mutual promoter repression, we will also integrate constructs in which the transcription termination signal of one of the two strands lies upstream of the opposing promoter (Fig 6.5d). These perturbations should allow us to identify the characteristics that are essential for the generation of a transcriptional memory.

The construction of an experimental system that allows testing of our model predictions is an essential but also challenging part of our analysis. If we indeed succeed to construct an antisense toggle switch, it can form a synthetic cellular memory unit that beyond helping us to understand biological phenomena could be useful also in biotechnological applications, e.g. to achieve sustained production of a target compound without the need for continuous induction.

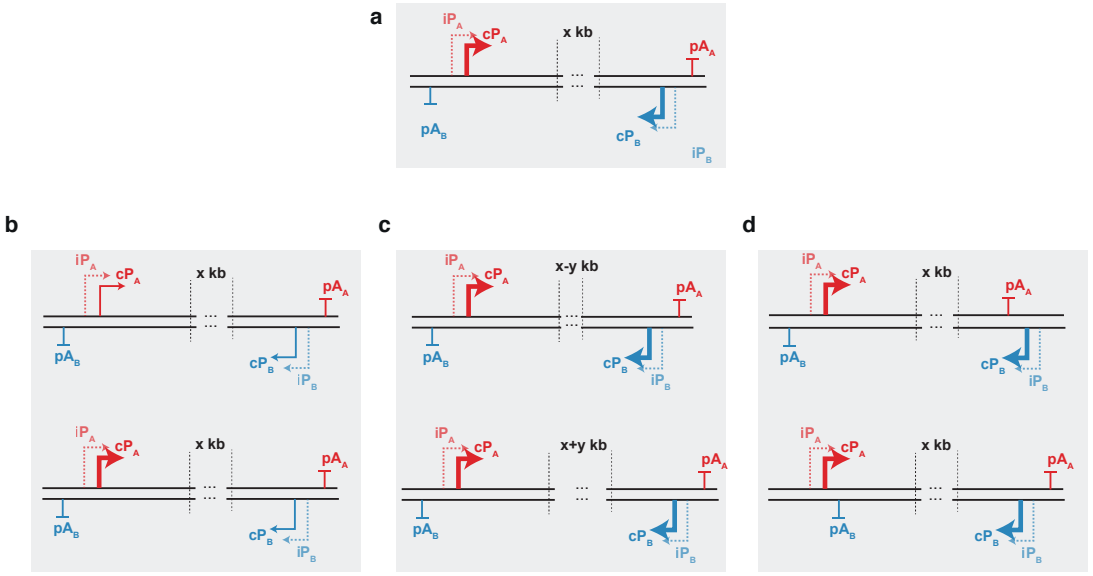

FIGURE 6.5: Experimental perturbation of critical parameters for transcriptional memory. a) Default construct predicted to display transcriptional memory with strong balanced promoters, long inter-promoter distance and both promoters residing within the overlap. b) Analyzing the impact of promoter strength using a construct with weak balanced promoters (top), or a construct with unbalanced promoters (bottom). c) Analyzing the impact of inter-promoter distance using a construct with decreased (top) or increased (bottom) length of the intronic sequence between promoters. d) Analyzing the importance of mutual promoter repression using constructs in which transcription from one of the two strands is terminated upstream of the opposing promoter.

Part III

Concluding remarks

This work has set out to identify the regulatory principles behind female-specific and monoallelic XCI in mammals and finds that this intriguing regulation can be ensured by two regulators only, a *trans*-acting XCI activator and a *cis*-acting XCI repressor. We propose that the regulatory network governing XCI in different species might be less diverse than previously assumed. Differences in Xist expression patterns between different species can arise from quantitative differences in reaction rates rather than qualitative differences in the network structure. We present experimental evidence for the existence of the negative feedback, by demonstrating for the first time that cells are able to revert biallelic to monoallelic Xist expression. Experimental evidence for the positive feedback and the dissection of its exact implementation remain to be provided. We however present a theoretical model that demonstrates that antisense transcription can stabilize alternative Xist expression states, and correctly predicts the phenotypes of several mutant lines. A central assumption of this model, the existence of transcriptional interference between Xist and Tsix, is validated experimentally.

Intrigued by the potential role of antisense transcription in setting up alternative Xist expression states, we then conducted a more detailed theoretical analysis of the potential for the generation of *cis*-encoded transcriptional memory by antisense loci. Our analysis predicts that antisense transcription can indeed stabilize alternative transcription states for several hours to days, a regulatory feature that can be useful in various biological contexts, for instance to filter out signal fluctuations or generate a short-term memory of previous stimulations.

This work demonstrates the power of interdisciplinary approaches that combine theoretical and experimental analyses in understanding fascinating biological phenomena, such as X-chromosome inactivation. Mathematical modeling allows us to formalize conceptual ideas and extract novel predictions that are then subjected to rigorous experimental testing.

Part IV

Materials and methods

7 Materials

Table 7.1: Antibodies

Antibody	Supplier	Number	Dilution
H3K27me3 antibody	Active Motif	39155	1/2500
goat anti-rabbit, labelled with Alexa555	Invitrogen	A-21428	1/2500

Table 7.2: PCR primer pairs

Primer name	forward	reverse	Use
VM151/VM152	TTCCAGCCATCACTTTGCGC	TGAGTGTGAGCTTTGGGGCA	XX vs XO (Length polymorphism Cast/B6)
ES585/ES595	AGGCACACCACCCCAGTGGA	AGTCAGAAGTTCAGTGTGAGACACCT	Genotyping Tx1072dT$^{-/-}$
AB15/AB16	TTCCCGCCATGTGATTATGC	ACGAACCACTGAGAATTGTTTTG	Genotyping Tx1072dT$^{-/-}$
VM207/BB005	GGTACCTAATAACTCGAGTGGCGATCC	GTCATGTGCACCTGCTCAGT	Genotyping TxΔXic_LPchr1
VM209/BB005	TCCATCACCTCAGACAACGC	GTCATGTGCACCTGCTCAGT	Genotyping TxΔXic_LPchr1
VM178/VM179	CATATGGACTACAAGGACGACGATGACAAGTCTATCGAACAAGCATGCGATATTTG	GTGGCGCCGCTGCCGCTAGCCCACCGTACTCGTCAATTC	Cloning of SP265
VM180/VM181	GGTGTCGTGACGCGGGATCCGGTACCGCCACCATGGCTGGAGACATGAGAGC	TCGATAGACTTGTCATCGTCGTCCTTGTAGTCCATATGGTCGACAGCTGTGGCAGG	Cloning of SP265
VM246/VM247	GAAGCACTCGAGCCAGTCACACCTCAGCTA	TGCTTAGGATCCAAGGATCTGCGGCACGCTGT	Cloning of SP271
BB002/BB003	TTACATGCAGGCATGCAAGCTTGGCGTAATC	CAGGCGCGCCAACTCGAATTTTGTAAAACAAACCGTGAAACGTC	Cloning of SP268
BB001/VM206	GTCTCATCATTTTGGCAAAGAATTCGCCACTATGGTGGTTTG	TCGAGTTATTAGGTACCTTCCTTTGCCCTCGGACG	Cloning of SP268
VM207/VM208	GGTACCTAATAACTCGAGTGGCGATCC	GCTTGCATGCCTGCATGTAAGGTTAGTGTGTCCAGC	Cloning of SP268
VM189/VM190	TCTAGCCGTCTCCCTGAATTTAAATGTTTATTGCAGCTTATAATGGTTAC	TCTAGCCGTCTCCCTCGGGGAAATGGATCCAGACATGATAAGATACATT	Cloning of SP273
VM191/VM192	TCTAGCCGTCTCGCGAGTTAACCACGCCACCTCGACATAC	TCTAGCCGTCTCGAATCCAGGCGATCTGACGGT	Cloning of SP273
VM197/VM198	TCTAGCCGTCTCCCCACGTGCCTGCAGGTCACGACACCTGAAATGG	TCTAGCCGTCTCCGATTAATTAAGTGCCCGTCAGTGGGCAGA	Cloning of SP273
VM201/VM202	TCTAGCCGTCTCCGTGGCTAGCGCCACCATGAAGATCAAGCTGTGCA	TCTAGCCGTCTCCCATATGTTATTAGGCCTGGCTGGGCAGCAT	Cloning of SP273
VM203/VM204	TCTAGCCGTCTCCTATGCACGTCCGTCGACGGATCCAGACATGATAAGATACATT	TCTAGCCGTCTCGTGCTGTGCGCATGTTTATTGCAGCTTATAATGGTTAC	Cloning of SP273
VM250/VM251	TGCTTACACGTGATGGGAGTGCAGGTGGAAAC	TGCTTAGCTAGCTTCCGGTTTTAGAAGCTCCAC	Cloning of SP279
VM264/VM265	TTCAGGTGTCGTGACCTGCAGGGGCAGGTTCGCTGTG	TCCACCTGCACTCCCATCACGTGATCCATGATGTTCACTTTCTTCTTGG	Cloning of SP302

Table 7.3: qPCR primer

qPCR	forward	reverse
Tsix spliced	CCGAGATATCCACGCATCTT	CGGGTTCTTGGTCGATGTAA
Xist	GGTTCTCTCTCCAGAAGCTAGGAA	TGGTAGATGGCATTGTGTATTATATGG
Rrm2	CCGAGCTGGAAAGTAAAGCG	ATGGGAAAGACAACGAAGCG
Arpo	TCCAGAGGCACCATTGAAATT	TCGCTGGCTCCCACCTT

Table 7.4: TruSeq Targeted RNA Expression assay amplicons

Gene Name	Transcript IDs	SNP position (mm10)	SNP ID (dbSNP)	Chromosome	Assay ID
Xist	NR_001570, NR_001463	10030062	rs29080493	X	10030062
Xist	NR_001570, NR_001463	10030417	rs31372548	X	10030417
Xist	NR_001570, NR_001463	10038162	rs31823151	X	10038162
Xist	NR_001463	10055737	rs29080913	X	10055737
Rrm2	NM_009104			12	7285426
Arpo	NM_007475			5	7285083
Fbxo28	NM_175127			1	7237479
Exoc1	NM_027270			5	7235075

Table 7.5: sgRNAs

Target	Sequence	Use
Tsix	GAACTCACTATATCGCCAAAG	Deletion of DxPas34
Tsix	GTACATAATGACCCGATCTC	Deletion of DxPas34
Chr1 neutral TAD	GAGCCCTGCGAAACACGCCA	Insertion of Landing Pad

Table 7.6: single-stranded repair oligos

Name	Target	Sequence	Use
LR160	Tsix	TTAAAGGGTCTCATGCATCCTAGCCTGGCCTCAA-ACTCACTATATCGCCACTCTGCTATCGTTGAAAT-TCGAAAATAATCTGAGTGCTTTGAAAACCCAT	Deletion of Tsix promoter and DxPas34

Table 7.7: Plasmids used in this study

Plasmid	Use	Source	Bacterial Resistence
JPF0446-BxBI-CAG-H (SP246)	Positive control vector expressing GFP for BxB1-mediated integrations into landing pad	Gift from Alain Bonny [299]	Kanamycin
pCas9-2A-mCherry (SP176)	sgRNA cloning for TX1072dT$^{-/-}$	Gift from Marc Bühler	Ampicillin
#178-pC2P-gRNA-chr1-new-knock-in (SP244)	Cas9 & sgRNA targeting chr1:73,641,436, mm10	Gift from Luca Giorgetti	Ampicillin
MTK0_017_-_BxBIattB_tag (SP248)	Backbone for Antisense construct	Gift from Alain Bonny [299]	Kanamycin
pCAG-NLS-HA-Bxb1 (SP225, Addgene 51271)	Expression of BxB1 for integrations into landing pad	Addgene 51271	Ampicillin
pLP1	Lentiviral packaging	Thermo Fisher Scientific	Ampicillin

Continued on next page

Table 7.7 – *Continued from previous page*

Plasmid	**Use**	**Source**	**Bacterial Resistence**
pLP2	Lentiviral packaging	Thermo Fisher Scientific	Ampicillin
VSVG	Lentiviral packaging	Thermo Fisher Scientific	Ampicillin
pLenti-Cas9- P2A-Puro (SP223)	Cloning of SP265	Addgene 110837	Ampicillin
pCAG-ERT2-PhoCl-Gal4VP16-PhoCl-ERT2 (SP228)	Cloning of SP265	Addgene 87693	Ampicillin
pLenti-PGK-EGFP-Blast (SP51)	Cloning of SP270 & SP271	Addgene 19069	Ampicillin
pTRETightBI-RY-0 (SP81)	Cloning of SP270 & SP273	Addgene 31463	Ampicillin
pSUPER-EF1-Puro-U6-sgRNA-UAS-mTK-mCherry (SP83)	Cloning of SP271	Cloned by Rutger Gjaltema	Ampicillin
#235-237-pUC19-Chr1-Ho-CuO-Cre-HyTK (SP242)	Cloning of SP268	Gift from Luca Giorgetti	Ampicillin
MTK0-057-BxBI-LP-on-h (SP250)	Cloning of SP268	Gift from Alain Bonny [299]	Kanamycin
pCas9-2A-mCherry (SP176)	Cloning of SP176_Tsix1 & SP176_Tsix2	Gift from Marc Bühler	Ampicillin
MTK0-017-BxBIattB-tag (SP248)	Cloning of SP273	Gift from Alain Bonny [299]	Kanamycin
pSLQ2817-pPB (SP102)	Cloning of SP273	Addgene 84239	Ampicillin
pU6-sgRNA-EF1Alpha-puro-T2A-BFP (SP125)	Cloning of SP273	Cloned by Liat Ravid	Ampicillin
MTK3b-004-mAzamiGreen (SP252)	Cloning of SP273	Gift from Alain Bonny [299]	Chloramphenicol
pBMN-FKBP(DD)-YFP (SP71)	Cloning of SP279	Addgene 31763	Ampicillin

Table 7.8: Stellaris RNA FISH probes

Xist exon Quasar 570	**Tsix 5' Fluorescein**	**Tsix 3' Quasar 570**	**Xist intron Quasar 670**	**Huwe1 intron Quasar 570**
acaaagattgggctgtcgag	cttgcaggtacttttgggac	gtactaccttgttagctact	ctttgacttttcagtgggtg	agaactgagctttaggagca
tcctgtacgacctaaatgtc	agtgagaacagagcaagggc	actctggctgtttagactac	gctgagaagagctagtgact	gaccagctatatagcagttg
ataccgcaccaagaacttga	aagatgcgtggatatctcgg	tcaagtaactcatccagtgc	tggcatactatgtgttgtgt	tactccttcactaagcacag
atgtgacacacaaagccctt	actatcccgcaagaatttca	tcatcacaacagcagttctc	tggcgattgggtacacaatg	gtaggtctttttgttctcat
gctaactcagacatgaagct	actgtcctgaaagcactttg	gctgctgagtgtttgatatg	atcgcttctcattttagtgt	ggggctcttgtagatactaa
gccatcttagacacattcaa	aggcattttaccttcatctc	ccgagtacttaggtcttttt	tcacacaggcagtacatttc	gccgttattgcaactttatc
ccataatgcaccaagtgtac	aactatcggttcgggactac	agctttctctagtgtgaagg	aatttcaaccccagtagact	ctcctgaaccttatcatagt
cttcgagtggcacaaggtag	attttagttattcctccgtt	ccaccattaaactctggcaa	cacttccctctttaagaagt	gaggccattattgtgagtat
aaccaacacttccacttagc	tagtctaaccacctgtaagg	tgacctaagccttctaatca	tggcgtacagctaagcataa	acccgtggacaattttcaag
ccatttctgacgagttacgt	gtatgttagctatctttcca	aatctttgtgtggcctttag	atattcccaattcctacaca	cttattctccctcaaaggat
accaacactttcacacttgt	atccatttgagccattagtc	caagtatctttgcttggtct	ttctgggcagaaatgcagta	accaattcattggactgtct
ccaccaagcaataatgcaca	gcattttagtaaagtctccc	tgtactgagggtgatgagtc	ttatccttctttgcactttt	agctgaactttaccaggaca
atcacagtctaattccatcc	cgtggcagacattttagtga	tgcaaggtggattgactgtg	tctgttgaaaagtgctggca	gagaaatgaatgaggggcca
aactcacgtccttatgggac	cgtggcagacattttagtga	caggaccaactgccagaaaa	gctagaccattttctacaca	ctgatgacatggtgcaggaa
acacacgtgaagtaccaagc	cgtggcagacattttagtga	cccgattttgtatattgcta	agtactagagggtcatttgc	agcatacaaggtatgggcaa
tgggcaggcacttcgaaaaa	ttttagtgatagcccagatc	gccaactgtctgcttaagaa	aacctgctgcgaaggaattt	tgttgtctagaggcattcat
tagaattgcaagcatgcgct	acattttagtgatagcccag	ggaagcttagcacttttctt	aatgctaagcctttggagtt	gccaagctttatattcaacc
cgacatcatccaacacttca	acattttagtgatagcccag	atattgcaggcaagtagctc	tattatcacctagagagcca	atagctgggagtgaggtatg
tttacatcgaccaagaaccc	ggcaggcattttagtgatag	attcttgcctacctttatta	caatggtcctagtttcctaa	tttttaaccctatatccctt
aatgtccttgaaatggcctt	taaggcagggattttagcga	ttattccaattccaactgct	ccttgcaaatgtttccttttt	actttggagttgtctctaca
atggaacgggctgagtttta	cattttagtcgacctctagg	gccttcaatctttaacttgc	agctaactgtcttagtgcac	tggaacagtggccatagata
agtagactagccactgaaca	ctatcccagacattttcatc	caggccaaaactaggtggat	aatggtgctaaggtgagacc	caggtaggacatgagactgg
tggctaaatcagaggccaag	attttagccaaccatctcta	tgggtaatgttaattacccc	cataggaggcaggctgaaag	cttccataagtggggtattt
agagaaccgcttgagatcag	gcttattgaacgctttgcat	atatggctacagtgtctcag	agtctgagtcctaatctcta	aagtatgcttcctgctattt
aaagtcaccacttataccca	tagactgcagtttattaccc	acaacaccagtttactcact	gtgaatagcagcagtaccag	accaggttctgtttgtatac
ttgtgatccactatgactct	acgcccttaactttcttttc	cccttctcttgacttaaagt	aactgtcagtgatagtgggt	atgtaattccctctcagatc
aaggtaggattctacctctt	attctagaccctgctacaag	cttttgcctctttgtttgaa	ctgctcacacaacttgatca	cggatgactgcattctctat
agtcatggggtagattttgg	agagtttttcggacacgagg	ctttctcaattccatcatgc	attggccctgattaatgatc	agcatgcttctttttcacta
aatcaaaatgcaaccccagc	ggtgaggtagtattgctgaa	ctcacactttatagtgttcc	aggtcatgtaggactaggtt	gtatatagctctctaggtca
aagcacactatcagacgtgt	cccattggtttttagcataa	caacaatgacccggatttcg	agtgtgtattttacagtgcc	tactttctcctatctggtta
tttgaagagtagctcggtgg	gaaaaccccactaaagcgga	tgctagtttcccaatgatat	acaagctcatgttccagatg	gtctcacactttgagcttag
tcctgacatccagtcataat	gctaccagtaatttttccac	actccaacttgagcaaatcc	ttatggggaccaaagatggc	cttggctgatgctacattta
aggctcttgctgacaagtaa	gcagacttagttactcgttc	gcagttctggatggtttatt	gggaactctttgtattccat	tgtaggcttggtatatgagc
ctgatagaggacacagagca	cttccaaggcttctcaaaga	tgggactcttaataggcaca	tgattactgacagtgagggc	gctcattcatttcttttcga
atctagatgccataaaggca	caacgataccagagatcggg	tgctctcttccttaaactga	gaagatggtgatggcgagtt	gggtcaaacttggtaatggc
agcactatttgctgagtctt	gggttttcaaagcactcaga	tactgtcaatgagcacaggg	atgtatacagttcaggcagc	agaaaacgggggagccaaga
tggacactgcattttagcac	gtcccagatagtagaaactt	acattattgcagaggttgct	ggacagcttggaagaaagct	atagtactgcttatgccttg
tctcagtcttataggctgag	tgagaattgttttgccttgc	tgtctggaatattttgtccc	acaaactcctaagtctgggt	ggatgtcctatgagaacact
tcagtatggaggggtatag	ttccgtgcctatagtaaata	cataccatcaaggctactga	attcacttgttctacagggt	ccccaaagtccagataatta
ggcactgcattttagcaata	ataaacactaccgggaggga	aaggttctgaaatgcctacc	ccaccacttaatgaaggtgt	ctattcttcaaggctctgtt
gaggggtatggggtataaa	ttgtttgctcgtctgaacac	ccttacacatatgacctgtg	gcaaaatccactacttccac	tttgtcctctaaccaaatcc
aattccatgaccctagaagt	actggattcattccttagtg	tgagcacttctcttgtcaat	gccacagggctttgtaaaat	aacaccagaaccccataaga
agggagtatgggtatacat	attgttaatttccctccttt	ttttgacaccagaggcatat	atacaaagacctggtctggt	ctgcttaatagccaagtagc
agcaataggaccgtatgcaa	ccaaatccgcatcaaaacca	ggagttggattttacatcct	gctaggttatatggcagtat	aaagcccttaagaagcctta
tctcagtcttataggctgag	tctttcttctcttccgaatt	aatgcagttagttgtccatc	cagaagaggttcactggact	gcctttgtaatcaatcaggt
cagtcttataggctgagtga	tttgtctgcctactaacaca	cgaagtgctcttcatttgga	ccagaggactaagtaagggg	agtaagggcaatgtgctaca
ggcactgcattttagcaata	ttaccttctaggggaatctt	aagaactactgtaccctctt	cacaaatagagccatcttcc	tagctgagaaggaacctctt
ctatctcagtcttataggct	ccctgccatttcaatgaaac	tcacatgtctaagcacttct	aaggcctttcatttcttttc	tgtatagtcttgtcattgcc

Table 7.9: Enzymes used in this study

Enzyme	Catalogue number	Company
HotStarTaq DNA polymerase	203203	Qiagen
Phusion High-Fidelity DNA polymerase	M0530L	NEB
T4 DNA Ligase	M0202S	NEB
AgeI-HF	R3552S	NEB
BamHI-HF	R3136S	NEB
BsaAI	R0531S	NEB
BsaI-HF	R3733S	NEB
BsmBI	R0580	NEB
EcoRI-HF	R3101S	NEB
Esp3I	R0734S	NEB
KpnI-HF	R3142S	NEB
NcoI-HF	R3193S	NEB
NdeI	R0111S	NEB
NheI-HF	R3131S	NEB
NIaIII	R0125S	NEB
PciI	R0655S	NEB
PmlI	R0532S	NEB
PstI	R3140S	NEB
SbfI-HF	R3642S	NEB
XhoI	R0146S	NEB

Table 7.10: Kits used in this study

Kit	Use	Catalogue number	Company
Direct-Zol RNA Miniprep Kit	RNA purification	R2070	Zymo Research
DNeasy Blood and Tissue kit	gDNA isolation	69504	Qiagen
Click-iT^{TM} EdU Alexa Fluor^{TM}647	EdU Proliferation Assay	C10340	Invitrogen
NucleoBond Xtra Midi Plus	Plasmid DNA purification	740412.50	Macherey-Nagel
NucleoSpin Gel and PCR Clean-up	PCR Clean-up	740609.50	Macherey-Nagel
Peqgold Plasmid Miniprep Kit	Plasmid DNA purification	732-2780	VWR
QIAquick Gel Extraction Kit	Gel Extraction	28704	Qiagen
TruSeq Targeted RNA Custom Kit	Allele-specific amplicon sequencing	RT-101-1001	Illumina
TruSeq Targeted RNA Index Kit	Allele-specific amplicon sequencing	RT-401-1001	Illumina
Turbo DNA free Kit	DNase digest	AM1907	Thermo Fisher Scientific
Gibson assembly cloning kit	Gibson cloning	E2611S	NEB

8 Experimental methods

8.1 Cell culture

8.1.1 Cell lines

Table 8.1: Cell lines

Cell line	Source	Genotype	Genetic Background	Parental Line	Antibiotic Resistance
Tx1072	Generated by Edda Schulz [58]	XX, rtTA in Rosa26, TetOP-Xist on B6	Cast x B6		Neomycin
TX1072dT$^{+/-}$ (clone 1C6)	Generated by Edda Schulz [165]	XX, rtTA in Rosa26, TetOP-Xist on B6, Dxpas34 deletion on Cast	Cast x B6	Tx1072	Neomycin
TX1072dT$^{-/-}$ (clones C9, C10, B6)	Generated in this study	XX, rtTA in Rosa26, TetOP-Xist on B6, Dxpas34 deletion on Cast and B6	Cast x B6	Tx1072	Neomycin
TXY	Gift from Anton Wutz [12]	XY, rtTA in Rosa26, TetOP-Xist	129/Ola	J1	/
TXYΔA	Gift from Anton Wutz [12]	XY, rtTA in Rosa26, TetOP-Xist, Deletion of Repeat A	129/Ola	J1	/
TxdXic_LPchr1 (clone C8, B8)	Generated in this study, based on [299]	XX, Deletion of Xic on B6, BxB1 Landing Pad on Chr1	Cast x B6	TxdXic A1 [92]	Neomycin, Hygromycin
TxdXic LPchr1_SP246	Generated in this study	XX, Deletion of Xic on B6, GFP (SP246) integrated in Landing Pad on Chr1	Cast x B6	TxdXic LPchr1 (clone C8)	Neomycin, Blasticidin

Table 8.1 contains a list of all cell lines used or generated in this study.

8.1.2 Cell culture conditions

All mESCs except TXY and TXYΔA were grown without feeder cells on gelatin-coated flasks (0.1%, Millipore) in serum containing ESC medium (DMEM (Sigma), 15% FBS (Pan Biotech), 0.1 mM β-Mercaptoethanol (Sigma), 1,000 $\frac{U}{ml}$ leukemia inhibitory factor (LIF, Millipore), supplemented with 2i (3 μM Gsk3 inhibitor CT-99021 (Axon Medchem), 1 μM MEK inhibitor PD0325901 (Axon Medchem)) where indicated. mESCs were passaged every second day to a density of $4 \cdot 10^4 \frac{cells}{cm^2}$. Medium was changed daily. Male TXY and TXYΔA inducible Xist lines were plated at a density of $3 \cdot 10^4 \frac{cells}{cm^2}$ on mitomycin C-inactivated mouse embryonic fibroblasts in ESC media containing 15% FBS (Pan Biotech), 0.1 mM β-mercaptoethanol (Sigma), 1,000 $\frac{U}{ml}$ LIF (Millipore). Differentiation was induced by (2i/)LIF withdrawal in DMEM supplemented with 10% FBS and 0.1 mM β-mercaptoethanol at a density of $4 \cdot 10^4 \frac{cells}{cm^2}$ in fibronectin (10 $\frac{\mu g}{ml}$ (Merck)) coated tissue culture plates. To induce Xist ectopically, the medium was supplemented with 1 $\frac{\mu g}{ml}$ doxycycline (Clontech). To induce Xist in undifferentiated cells, cells were plated at a density of $1 \cdot 10^5 \frac{cells}{cm^2}$ two days before collection and treated with 1 $\frac{\mu g}{ml}$ doxycycline. To assess the effect of Xist induction on Tsix transcription, male TXY and TXYΔA lines where induced for 24 h with 2 $\frac{\mu g}{ml}$ doxycycline one day after seeding.

HEK293T cells were cultured in DMEM supplemented with 10% FBS and passaged every three days.

8.1.3 Mice

All animal experiments were performed by Ikuhiro Okamoto as described in [165].

8.1.4 Lentiviral transduction

For lentiviral transductions, a third generation transfer system was used, which consists of an envelope plasmid (VSVG), two plasmids that both contain part of the packaging system (pLP1, pLP2) and the actual transfer plasmid that contains the DNA of interest flanked by long terminal repeat (LTR) sequences allowing the integration into the host genome (Thermo Fisher Scientific). To package DNA constructs into lentiviral particles, $1 \cdot 10^6$ HEK293T cells were seeded into a well of a 6-well plate. After one day, they were transfected with the lentiviral plasmids: 1.2 μg pLP1, 0.6 μg pLP2, 0.4 μg VSVG and 2 μg of the transfer plasmid using lipofectamine 2000 (Thermo Fisher Scientific). HEK293T supernatant containing the viral particles was harvested after 48h. For lentiviral transduction of the mESCs, $0.2 \cdot 10^6$ mESCs were seeded in a well of a 12-well plate and transduced the following day with 500 μl of viral supernatant in medium supplemented with $8 \frac{ng}{\mu l}$ polybrene (Merck) to increase viral infection efficiency, by neutralizing the charge repulsion between virus and cell surface. Two days after transduction, antibiotic selection (Puromycin (1 μg/ml, Sigma) or Blasticidin (5 μg/ml, Roth)) was started and maintained for at least three passages.

8.1.5 Genome editing

For genome editing targeted double strand breaks at the genomic site of interest were introduced using the CRISPR/Cas9 technology, which allows targeting of the Cas9 nuclease to the site of interest via a single guide RNA (sgRNA). Repair templates with homology arms to both sites of the break were supplied to facilitate repair of the double strand break via homology directed repair (HDR).

8.1.5.1 Tx1072dT$^{-/-}$

Tx1072dT$^{-/-}$ lines were generated by deleting a 4 kb region encompassing the DxPas34 repeat and the major Tsix promoter (mm10 chrX:103,446,011-103,450,162). To generate this line, two guides (Tsix1 and Tsix2) flanking the region to be deleted were cloned into the pCas9-2A-mCherry vector, and transfected together with a single-stranded oligo deoxynucleotide (ssODN) repair template with 50 bp homology flanking each cut site (LR160). The sequences of the sgRNAs and ssODN are provided in Tables 7.5 and 7.6, respectively. $0.5 \cdot 10^6$ cells were seeded in a well of a 6-well plate and reverse transfected with 1.25 μg of each guide plasmid and 10pmol of the ssODN repair oligo using Lipofectamine 3000 (Thermo Fisher Scientific) according to the manufacturer's instructions. Two days post transfection mCherry positive cells were sorted and seeded for clone picking at a density of 15 $\frac{cells}{cm^2}$ into 10 cm plates. After 10 days single clones were transferred to one well of a 96-well plate and expanded. Genomic DNA was extracted as described in section 8.2.4 and screened for the presence of the deletion with primers ES585/ES595, the presence of the WT with primers AB15/AB16, and the presence of both X chromosomes (Cast and B6) with primers VM151/VM152 using HotStarTaq polymerase (Qiagen) with a Tm=60°C and 30 cycles (see section 8.2.4). Fig 8.1 summarizes the genotyping.

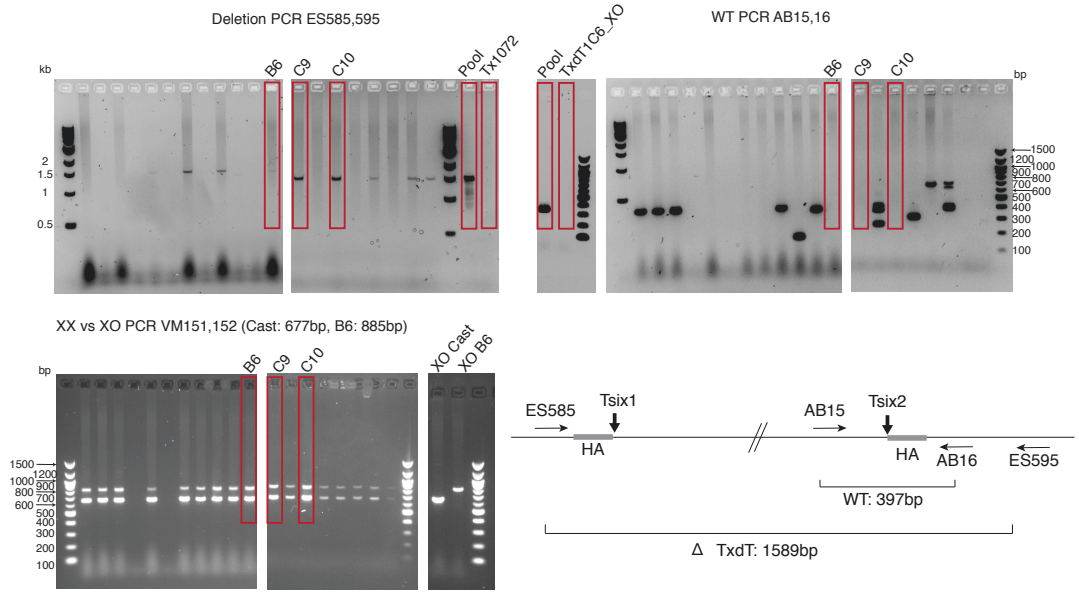

FIGURE 8.1: Genotyping of the Tx1072dT$^{-/-}$ clones. In the genotyping PCRs controls and chosen clones B6_3, C9_3 and C10_3 are highlighted in red (Pool: Pool of transfected cells; TxdT1C6_XO: Cast XO line with Tsix deletion). In the scheme, position of the Cas9 cut sites (vertical arrows), of the homology arms in the repair template (grey bars) and of the primers (horizontal arrows) used for genotyping are shown together with the expected sizes for the PCR products.

8.1.5.2 TxdXic_LPchr1

TxdXic_LPchr1 lines were generated by B. Boesen, supervised by V. Mutzel, using CRISPR/-Cas9 mediated homologous recombination. One guide targeting the insertion point (chr1:73,641,436, mm10) $0.5 \cdot 10^6$ cells were seeded in a well of a 6-well plate and reverse transfected with 1.5 μg of the guide plasmid and 1 μg of the repair template using Lipofectamine 3000 according to the manufacturer's instructions. Two days post transfection Hygromycin selection (200 μg/ml, VWR) was started, and one day later cells were passaged to a 10cm plate at a density of 15 $\frac{cells}{cm^2}$. 10 days later, single clones were transferred to one well of a 96-well plate and expanded. Genomic DNA was extracted as described in section 8.2.4 and screened for the presence of the insertion with primers VM207/BB005 and the presence of the WT with primers VM209/BB005 using HotstartTaq polymerase (Qiagen) with a Tm=55°C and 30 cycles (see section 8.2.4). To determine on which allele the landing pad was inserted, WT and MUT PCR bands were Sanger sequenced with BB009 and the SNP at position chr1:73642192 (mm10, Cast C) was used to assign WT and MUT allele. Fig 8.2 summarizes the genotyping.

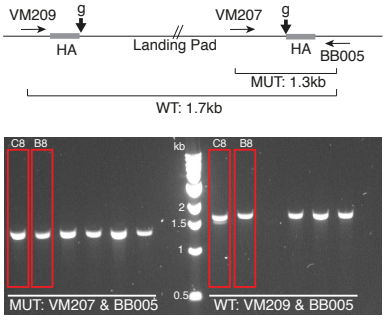

FIGURE 8.2: Genotyping of the TxdXic_LPchr1 clones. In the genotyping PCRs chosen clones C8 and B8 are highlighted in red. In the scheme, position of the Cas9 cut site (vertical arrows), of the homology arms in the repair template (grey bars) and of the primers (horizontal arrows) used for genotyping are shown together with the expected sizes for the PCR products.

8.1.6 BxB1-mediated integrations into landing pad

For BxB1-mediated integrations into the landing pad, $0.5 \cdot 10^6$ cells were seeded in a 6-well and reverse transfected with 1.5 μg of the transfer plasmid (e.g. SP246) and 1.5 μg of the BxB1-encoding plasmid (SP225) using Lipofectamine 3000 according to the manufacturer's instructions. Two days post transfection blasticidin selection (5 μg/ml, Roth) was started and maintained until control cells had died with daily medium changes. Then, reporter expression was assessed using flow cytometry.

8.1.7 Karyotyping

To ensure that cell lines had a normal karyotype, double digest genotyping-by-sequencing (ddGBS) was performed by Oriana Genolet and Ilona Dunkel as described here[1]. In short, gDNA was digested with NIaIII and PstI (NEB), followed by ligation of oligonucleotide adapters complementary to the generated overhangs. gDNA fragments were then PCR amplified using primers complementary to the ligated adaptors, and subsequently sequenced using next-generation sequencing [300].

8.2 Molecular biology methods

Standard molecular biology procedures, such as polymerase chain reaction (PCR) and gel electrophoresis, were conducted according to [301]. Table 8.2 summarizes the plasmids generated in this study together with the method used to generate them. After plasmids were generated, they were transformed into chemically competent bacteria (Stellar (Takara) or NEB stable (NEB)) according to the manufacturer's instructions and plated on agar plates with the respective antibiotics (Ampicillin: 100 μg/ml (Sigma); Kanamycin: 50 μg/ml (Sigma)). The following day, single colonies were inoculated in liquid medium with the respective antibiotics, and mini plasmid preparations were done using the PeqGold Plasmid Miniprep kit (VWR) following the manufacturer's instructions. After verifying plasmid identity, a colony was inoculated into 100 ml liquid medium with the respective antibiotics and cultured overnight at 37 °C and 200 rpm.

[1] https://palmerlab.org/protocols-data/

Midi plasmid preparations were done using the NucleoBond Xtra Midi Plus kit (Macherey-Nagel) following the manufacturer's instructions.

Table 8.2: Plasmids generated in this study. SP numbers as unique identifiers are also given.

Plasmid	**Use**	**Insert source**	**Original plasmid**	**Cloning method**	**Verification strategy**
pLenti-ERT2-FLAG-Gal4-NLS-VP16-P2A-Puro (SP265) (by B. Boesen)	Lentiviral integration of ERT2-Gal4-VP16	**(1)** Gal4-SV40NLS-VP16: amplified from pCAG-ERT2-PhoCl- Gal4VP16-PhoCl-ERT2 (SP228, Addgene 87693) with VM178& VM179 using Phusion with a Tm of 63°C; **(2)** ERT2 amplified from pCAG-ERT2-PhoCl-Gal4VP16-PhoCl-ERT2 (SP228, Addgene 87693) with VM180& VM181 using Phusion with a Tm of 63°C	pLenti-Cas9-P2A-Puro (SP223, Addgene 110837) NheI & BamHI digested	Gibson Assembly	Restriction digest with NdeI, NheI, KpnI followed by Sanger sequencing
pLenti-TetO-EGFP-Pgk-Blasti (SP270) (by B. Boesen)	Lentiviral integration of Dox inducible GFP	TetO-mCMV: excised from pTRETightBI-RY-0 (SP81) with XhoI & BamHI	pLenti-PGK-EGFP-Blast (SP51) XhoI & BamHI digested	Restriction cloning	Restriction digest with BamHI & XhoI followed by Sanger sequencing
pLenti-UAS-EGFP-Pgk-Blasti (SP271) (by B. Boesen)	Lentiviral integration of 4-OHT inducible GFP	UAS-mPgk: amplified from pSUPER-EF1-Puro-U6-sgRNA-UAS-mTK-mCherry (SP83) with VM246 & VM247 using Phusion with Tm of 61°C	pLenti-PGK-EGFP-Blast (SP51) XhoI & BamHI digested	Restriction cloning	Restriction digest with BamHI & XhoI followed by Sanger sequencing
Landing Pad vector (SP268) (by B. Boesen)	HDR template for TxΔXic-LPchr1 generation	**(1)** HS4-HS4-LoxP-pCAG: excised from MTK0-057-BxBI-LP-on-h (SP250) with EcoRI; **(2)** HygroR: amplified from SP250 with BB001 & VM206 using Phusion with Tm of 64°C; **(3)** SV40pA-InsulatorA1: amplified from SP250 with VM207 & VM208 using Phusion with Tm of 61°C	#235-237-pUC19-Chr1-Ho-CuO-Cre-HyTK (SP242) amplified with BB002 & BB003 using Phusion with Tm of 63°C	Gibson Assembly	Restriction digest with NcoI followed by Sanger sequencing

Continued on next page

Table 8.2 – *Continued from previous page*

Plasmid	**Use**	**Insert source**	**Original plasmid**	**Cloning method**	**Verification strategy**
pCas9-2A-mCherry-Tsix1 (SP176_Tsix1) (by V. Mutzel)	Cas9 and Tsix1 sgRNA for Tx1072dT$^{-/-}$ generation	annealed sgRNA oligos LR156 & LR157	pCas9-2A-mCherry (SP176) BsaI digested	Restriction cloning	Restriction digest with BsaI & AgeI followed by Sanger sequencing
pCas9-2A-mCherry-Tsix2 (SP176_Tsix2) (by V. Mutzel)	Cas9 and Tsix2 sgRNA for Tx1072dT$^{-/-}$ generation	annealed sgRNA oligos LR158 & LR159	pCas9-2A-mCherry (SP176) BsaI digested	Restriction cloning	Restriction digest with BsaI & AgeI followed by Sanger sequencing
AS construct step1 (SV40pA$_{AS}$-TetO$_S$-EF1a$_S$-mAzami$_S$-SV40pA$_S$, SP273) (by B. Boesen)	Intermediate step1 AS construct	**(1)** SV40pA_AS: amplified from SP102 with VM189& VM190 using Phusion with Tm of 60°C; **(2)** TetO_mCMV_S: amplified from SP81 with VM191& VM192 using Phusion with Tm of 63°C; **(3)** EF1a_S: amplified from SP125 with VM197& VM198 using Phusion with Tm of 61°C; **(4)** mAzami_S: amplified from SP252 with VM201& VM202 using Phusion with Tm of 61°C; **(5)** SV40pA_S: amplified from SP102 with VM203& VM204 using Phusion with Tm of 60°C	MTK0-017-BxBIattB-tag (SP248) BsmBI digested	Golden Gate cloning with BsmBI	Restriction digest with BamHI and BsaA1 followed by Sanger sequencing
AS construct step2 (SV40pA$_{AS}$-TetO$_S$-EF1a$_S$-FKBP$_S$-mAzami$_S$-SV40pA$_S$, SP279) (by B. Boesen)	Intermediate step2 AS construct	FKBP: amplified from SP71 with VM250& VM251 using Phusion with Tm of 72°C	SP273 NheI & PmlI digested	Restriction cloning	Restriction digest with BamHI & BsaAI followed by Sanger sequencing

Continued on next page

Table 8.2 – *Continued from previous page*

Plasmid	**Use**	**Insert source**	**Original plasmid**	**Cloning method**	**Verification strategy**
AS construct step3 (SV40pA$_{AS}$-TetO$_S$-EF1a$_S$-E1$_S$-I15'$_S$-mScarlet$_{AS}$-FKBP$_{AS}$-E2$_{AS}$-I13'$_{AS}$-I13'$_S$-E2$_S$-FKBP$_S$-mAzami$_S$-SV40pA$_S$, SP302) (by V. Mutzel)	Intermediate step3 AS construct	gBlock F4-7 ordered from IDT and amplified with VM264 & VM265 using Phusion with Tm of 67°C	SP279 PmlI & SbfI digested	Gibson Assembly	Restriction digest with PmlI & SbfI followed by Sanger sequencing
AS construct step 4 (SV40pA$_{AS}$-TetO$_S$-EF1a$_S$-E1$_S$-I15'$_S$-mScarlet$_{AS}$-FKBP$_{AS}$-E2_5'$_{AS}$-I1_3'$_{AS}$-I1_3'$_S$-E2_5'$_S$-FKBP$_S$-mAzami$_S$-E2_3'$_S$-I2_5'$_S$-I1_5'$_{AS}$-E1$_{AS}$-EF1a$_{AS}$-UAS$_{AS}$-SV40pA$_S$, SP348) (by Genscript)	Final AS construct	synthesis by Genscript	SP302	Cloned by Genscript	

8.2.1 Restriction enzyme cloning

For conventional restriction enzyme cloning, backbone and insert were digested overnight with respective enzyme(s) in recommended buffer and temperature. Digests were then run on an agarose gel and desired fragments were excised and gel purified using the QIAquick Gel Extraction Kit according to the manufacturer's instructions. Ligation was performed using T4 DNA Ligase (NEB) with $\sim$ 100ng total DNA combining digested backbone and insert in a 1:3 molar ratio.

8.2.2 Golden Gate cloning

For Golden Gate cloning, a 10 μl reaction was set up with 7.5 U Esp3I, 100 U T4 ligase, 1 mM ATP and equal molar amounts of backbone and each fragment in 1x CutSmart buffer (NEB). 100ng of backbone were used. The following cycling conditions were used for the Golden Gate reaction:

- 30 cycles of:
  5 min at 37°C
  5 min at 20°C
- 20 min 65°C

8.2.3 sgRNA design and cloning

sgRNAs were designed using the Zhang lab's CRISPR-Cas online tool (http://crispr.mit.edu:8079/). Off target scores based on off target prediction [302] were compared between candidate sgRNAs and the top scoring sgRNAs were chosen. sgRNA off-targets had to have at least 3 mismatches. For cloning of the guides into pCas9-2A-mCherry, two complementary oligos containing the guide sequence and BsaI homologous overhangs (oligo1: 5'caccgNNNNNNNNNNNNNNNNNNNN3'; oligo2: 5'aaacNNNNNNNNNNNNNNNNNNNNC3') were annealed and ligated into the BsaI linearized pCas9-2A-mCherry backbone, which contains the Cas9 endonuclease coupled to mCherry and a chimeric tracrRNA, allowing for sorting of cells transiently expressing Cas9.
Cloning was performed according to the Zhang lab protocol[2] except that dephosphorylation of backbone and phosphorylation of guides were omitted. The pCas9-2A-mCherry-sgRNA plasmids were transformed in Stellar competent E.coli (Takara). After Plasmid DNA isolation, correct sgRNA sequence was validated by Sanger sequencing. Correct plasmids were then produced in 100ml cultures and re-sequenced. An overview of all sgRNAs used in this study can be found in Table 7.5.

8.2.4 Genotyping

To genotype single clones, genomic DNA was isolated as follows: ESC clones were grown in 96-well plates until over-confluent. Cells were washed once with PBS, then 50 μl Bradley lysis buffer (10 mM Tris-HCl, 10 mM EDTA, 0.5% SDS, 10 mM NaCl) with freshly added proteinase K (1 mg/ml) were added to each well. The plate was sealed with parafilm, put into a humidified chamber and incubated at 55 °C overnight. The next day, 150 μl freshly made, ice-cold EtOH/NaCl mix (100% EtOH, 75 mM NaCl) were added per well to precipitate the DNA. The plate was incubated for 30 min at RT, and then centrifuged for 30 min at 4000 rpm. The plate was inverted to remove the liquid. The DNA was then washed twice by adding 150 μl ice-cold 70% EtOH, and centrifuging for 15 min at 4000 rpm. The pellet was air-dried by incubating the plate at 45 °C without lid for 10 min. 200 μl H2O were added per well, and the plate was

[2] https://www.addgene.org/crispr/zhang/

incubated at 37 °C for 2 h to resuspend the DNA. Unless indicated otherwise, 1 μl of this DNA extraction were used as template for the genotyping PCRs.

Single clones were genotyped by standard PCR using Taq Hotstart polymerase (Qiagen). Reagents were pipetted into a 96 well plate and DNA was amplified in a Thermocycler. PCR reaction mix for 25 μl reaction:

- 2.5 μl 10xBf
- 0.5 μl dNTPs (10 mM)
- 0.5 μl Forward Primer (10 μM)
- 0.5 μl Reverse Primer (10 μM)
- 0.125 μl Taq Polymerase
- 1 μl template DNA

PCR program:

- 15 min at 95°C
- 30 cycles of:
  1 min at 94°C
  1 min at Tm
  1 min/kb at 72°C
- 10 min at 72°C
- 4°C hold

The PCR products were analyzed by loading 10 μl of PCR product on a 1% agarose gel. Primer sequences for genotyping are listed in Table 7.2.

8.3 Flow cytometry

Integration of the EGFP reporter (Ex_{max}: 488nm, Em_{max}: 507nm) into the landing pad was assessed by flow cytometry. Cells transfected with the pCas9-2A-mCherry-sgRNA plasmids and transiently expressing mCherry (Ex_{max}: 587nm, Em_{max}: 610nm) were measured and sorted by FACS. For this, cells were resuspended to a density of $1 \cdot 10^7$ $\frac{cells}{ml}$ in sorting buffer (1x PBS (Sigma), 10%FBS, 1 mM EDTA). Cells were measured and/or sorted using the BD FACS Aria Fusion. The sideward and forward scatter areas were used to identify live mESCs, the height and width of the sideward and forward scatter signal were used to exclude duplets. FCS files were analyzed with the FlowJo^{TM} 10 Software (BD Biosciences).

8.4 RNA FISH

RNA FISH is an *in situ* hybridization method that allows to visualize target RNAs using flourescently labeled oligonucleotide probes that are complementary to the target RNA sequence. Different types of probes can be used for detection of target RNAs. Chromosomal fragments of interest can be cloned into plasmids or bacterial artificial chromosomes (BACs), and can then be labeled with fluorescent dyes through Nick-translation. To obtain strand-specific information, short single-stranded custom-designed DNA oligonucleotides complementary to various regions of the target RNA can be synthesized, each fluorescently labeled. These probes usually provide enough sensitivity to detect individual RNA molecules within single cells, so that the RNA of interest can be visualized, providing spatial information (e.g. which allele), and quantified. This technique can thus yield quantitative expression measurements at single cell, single allele resolution.

8.4.1 RNA FISH on mESCs

RNA FISH on mESCs was performed as described previously ([303]). In brief, cells were singularized with Accutase (Life Technologies) and attached to coverslips (Marienfeld) coated with Poly-L-Lysin (Sigma) for 5 min. Cells were then fixed with 3% Paraformalydehyde (Roth) in 1xPBS for 10 min at RT. After three washes with 1xPBS, cells were permeabilized for 5 min on ice (Permeabilization solution: 1xPBS, 2 mM Ribonucleoside Vanadyl complex (NEB), 0.5% TritonX-100 (Sigma)). Following three washes with 70% EtOH, coverslips were preserved at -20 °C in 70% EtOH for subsequent use in RNA FISH.

For RNA FISH combining Huwe1 BAC probes and custom-designed Xist probes (Roche), the fixed cells were dehydrated through an ethanol series (80%, 95%, 100% twice) and air-dried until no more ethanol was visible. To detect Huwe1, a BAC probe spanning the respective genomic region (RP24-157H12) was fluorescently labeled by nick translation (Abbot) using dUTP-Atto550 (Jena Bioscience). Per coverslip, 60 ng probe was ethanol precipitated with Cot1 repeats, resuspended in formamide, denatured (10 min 75 °C) and competed for 1 h at 37 °C. Xist was detected with a custom-designed strand-specific probe set of ∼75-bp-long oligonucleotides targeting all exons and labeled with Alexa488 (Roche). Both probes were co-hybridized in FISH hybrization buffer (50% formamide, 20% dextran sulfate, 2x SSC, 1 μg/μl BSA, 10 mM vanadyl-ribonucleoside) overnight. To reduce background, three 7 min washes were carried out at 42 °C in 50% formamide in 2x SSC at pH 7.2, followed by three 5 min washes in 2x SSC. Nuclei were stained with 0.2 mg/ml DAPI (4',6-Diamidin-2-phenylindol), and mounted in medium consisting of 90% glycerol, 0.1x PBS, 0.1% p-phenylenediamine at pH 9 (Sigma).

For RNA FISH with Stellaris FISH probes (Biosearch Technologies; probe details see Table 7.8), cells were prehybridized in wash buffer (2x SSC (Sigma), 10% formamide (Sigma)) for 5 min, then hybridized overnight at 37 °C (Hybridisation buffer: 125 nM of each FISH probe, 2x SSC, 10% formamide, 10% dextran sulfate (Sigma)). Cells were washed twice with wash buffer for 30 min. Then DNA was stained with 0.2 $\frac{mg}{ml}$ DAPI (Sigma) in wash buffer for 3 min. After another wash, cells were mounted on slides using Vectashield mounting medium.

Until imaging, slides were kept at -20 °C. Z-stacks were acquired using a wide-field Z1 observer (Zeiss) microscope equipped with a x100 objective (voxel size 88 x 88 x 200 nm).

8.4.2 Quantification of RNA FISH signals

For quantitative RNA FISH the Stellaris Xist and Tsix probes were used (Table 7.8). Cells were adsorbed, fixed, hybridized and imaged as described above. Quantification of nascent RNA signals was performed as described in [304], with analysis script provided by Luca Giorgetti. In brief, the fluorescence background of each z plane was generated by morphologically opening the image with a circular structuring element with a diameter of 5 pixels (440 nm), and subtracted from the original image. A region of interest (ROI) of constant volume (30 x 30 x 6 pixels = 2.6 x 2.6 x 1.2 μm) was selected around each transcription site. To reduce residual high-frequency fluorescence background, the average pixel intensity was measured in a 3-voxel-thick frame adjacent to the border of the ROI, and further subtracted. The integrated intensity of the fluorescent signal was then measured within the whole ROI. Integrated intensities of approximately 500 random nuclear background ROIs were used to define a threshold (mean + 5 s.d. after removing top 1% as outliers) to classify transcribed versus non-transcribed loci.

8.4.3 RNA FISH of epiblast cells from E5.0 embryos

RNA FISH on E5.0 epiblast cells was performed by Ikuhiro Okamoto as described here [165].

8.5 Immunofluorescence combined with RNA FISH

For immunofluorescence staining, cells were differentiated on fibronectin coated cover slips (18 mm, Marienfeld) at a density of $2 \cdot 10^4 \frac{cells}{cm^2}$. Cells were fixed and permeabilized as described above. They were then incubated with the H3K27me3 antibody (Active Motif 39155, 0.4 $\frac{\mu g}{ml}$) in PBS for 1 h at room temperature, then washed three times for 10 min with PBS, followed by a 1 h incubation with an Alexa-555 labelled Goat anti-rabbit antibody (Invitrogen A-21428, 0.8 $\frac{\mu g}{ml}$). After three washes, the cells were fixed again with 3% paraformaldehyde in PBS for 10 min at room temperature, followed by three short washes with PBS and two washes with 2x SSC. Hybridization was then performed as described above. Details on the antibodies used are found in Table 7.1. Immunofluorescence together with RNA FISH was performed by Ilona Dunkel.

8.6 EdU staining combined with RNA FISH

Cells were differentiated on fibronectin-coated cover slips (18 mm, Marienfeld) at a density of $2 \cdot 10^4 \frac{cells}{cm^2}$ and were treated with 7.5 μM EdU (Component A from Click-iT EdU Imaging kit, Invitrogen) for 2 h before collection. Cells were fixed and permeabilized as described in section 8.4, except that fixation and permeabilization were carried out at room temperature for 15

and 20 min, respectively. EdU staining with Alexa Fluor 647 was performed according to the manufacturer's recommendations, followed by RNA FISH for Xist using the Roche probe as described above. EdU staining together with RNA FISH was performed by Ilona Dunkel.

8.7 RNA extraction, reverse transcription, qPCR

Quantitative PCR (qPCR) allows to estimate RNA abundance by reverse transcription followed by a PCR reaction in which DNA amplification is monitored in real time through the use of DNA-binding fluorescent dyes, that specifically bind double-stranded DNA, such as SYBR Green.

For RNA extraction cells were washed with ice-cold PBS and lysed directly in the plate by adding Trizol (Invitrogen). RNA was isolated using the Direct-Zol RNA Miniprep Kit (Zymo Research) following the manufacturer's instructions. For qPCR, 1 μg RNA was reverse transcribed using Superscript III Reverse Transcriptase (Invitrogen) with random hexamer primers (Thermo Fisher Scientific). Expression levels were quantified using 2x SYBR Green Master Mix (Thermo Fisher Scientific). Expression levels were normalized to Rrm2 and Arpo. Primer sequences are given in Table 7.3.

8.8 Allele-specific amplicon sequencing

RNA was extracted using the Direct-zol RNA MiniPrep kit (Zymo Research) and DNase digest was performed using a Turbo DNA free kit (Thermo Fisher Scientific). The TruSeq Targeted RNA Expression assay (Illumina) was used according to the manufacturer's recommendations and the samples were sequenced on a HiSeq2500. For the quantification of reference genes (Rrm2, Rplp0, Fbxo28, Exoc1) 50 bp reads were aligned to the mouse reference genome (mm10) allowing two mismatches using the STAR aligner [305], and the reads covering each amplicon were counted with Bedtools multicov [306]. For allele-specific quantification, reads were aligned to either the B6 (reference) or Cast genomes with no mismatches and reads covering the SNPs were counted with Bedtool multicov. Reads for four amplicons within Xist exons containing SNPs were normalized to the geometric mean of the reference genes. The fold change of the doxycycline treated sample relative to the corresponding control sample was then calculated for each Xist amplicon. We then tested to see whether the mean log2 fold change of the four amplicons was significantly different from 0 ($P < 0.05$) using a one-sample t-test. Details on the amplicons are given in Table 7.4. Amplicon sequencing was performed by Ilona Dunkel and Norbert Mages, analysis was performed by Edda Schulz.

9 Computational methods

The code and simulations used in this study are available at `https://github.com/verenamutzel/XCI_model` or upon request.

9.1 ODE simulations

9.1.1 Model description

Alternative regulatory networks were formulated as ordinary differential equation (ODE) models and analyzed for their ability to stabilize monoallelic and female-specific Xist expression. Regulation of Xist by an activator or repressor and silencing of the regulator by Xist were formulated as Hill functions with a Hill coefficient n and a threshold K. Equations of the single-regulator models can be found in Table 2.1. Equations of the two-regulator models can be found in Table 2.2.

9.1.2 Simulating maintenance of monoallelic expression

To identify the networks that can maintain monoallelic Xist expression, a cell with one active (Xa) and one inactive (Xi) X chromosome was simulated with the initial conditions depicted in Table 9.1.

Each network was simulated with at least 10,000 parameter sets. Parameter values were randomly drawn from a uniform distribution between 1 and 5 for the Hill coefficients n, and from a logarithmic distribution between 0.01 and 10 for the threshold parameters K. Each model was integrated from 0 h to 100 h using the Matlab ODE solver *ode23tb*. To ensure that the steady state was reached, the state of the system after 100 h was then used as an initial guess in the Matlab non-linear equation solver *fsolve* to solve for the steady state numerically. If at the steady state $\text{Xist(Xi)} > 10 \cdot \text{Xist(Xa)}$ a parameter set was classified as monoallelic.

Table 9.1: Initial conditions for XaXi maintenance simulation

Xist x1 (Xi)	**Xist x2 (Xa)**	**Regulator r1 (Xi)**	**Regulator r2 (Xa)**
1	0.01	0.01 (silenced) 1 (escaping)	1

9.1.3 Simulating biallelic expression and male cells

To identify the networks that could destabilize biallelic Xist expression in female cells and prevent Xist expression in male cells, we performed three additional simulations, that initiated female cells from the XiXi state (Simulation 3), and male cells from the Xa state (Simulation 2) or the Xi state (Simulation 4). To simulate male cells, the models were reduced to containing a single X chromosome with one copy of Xist and each regulator (Table 9.2). For all parameter sets that were classified as monoallelic in section 9.1.2 the three additional simulations were performed with initial conditions summarized in Table 9.3. Based on the steady state on the Xi ($Xist^{high}$ state) in the monoallelic simulation in section 9.1.2, stability of XiXi, Xi or Xa in all parameter sets was classified as summarized in Table 9.4.

Table 9.2: ODE models of male cells

	Chr. X1	
Xist	$\frac{dx_1}{dt} = f(r) - x_1$	(4)
cis regulator r	$f(r) = a + b\frac{r_1^n}{r_1^n + K^n}$	
trans regulator r	$f(r) = a + b\frac{(0.5 \cdot r_1)^n}{(0.5 \cdot r_1)^n + K^n}$	
	XA: a=0, b=1 XR: a=1,b=-1	
Regulator		(5)
silenced	$\frac{dr_1}{dt} = 1 - \frac{x_1^n}{x_1^n + K^n} - r_1$	
escaping	$\frac{dr_1}{dt} = 1 - r_1$	

Table 9.3: Initial conditions for XiXi, Xi and Xa simulations

Simulation	**Xist x1**	**Xist x2**	**Regulator r1**	**Regulator r2**
XiXi (female)	0.99	1	0.01 (silenced) 1 (escaping)	0.01 (silenced) 1 (escaping)
Xi (male)	1		0.01 (silenced) 1 (escaping)	
Xa (male)	0.01		1	

Table 9.4: Stability classification criteria

Simulation	**Classification**	**Rule**
XiXi	XiXi unstable	$x_1 < 10 \cdot Xist^{high}$ OR $x_2 < 10 \cdot Xist^{high}$
Xi	Xi unstable	$x_1 < 10 \cdot Xist^{high}$
Xa	Xa stable	$x_1 < 10 \cdot Xist^{high}$

9.2 Stochastic cXR-tXA model

9.2.1 Model description

Since a stochastic symmetry breaking event is required to induce the transition from the symmetric XaXa to the asymmetric XaXi state, Xist upregulation was simulated stochastically. Table 2.4 in section 2.2 summarizes the adapted formulation of the model for the stochastic simulation.

Xist-dependent silencing is known to occur with a delay of hours or days after Xist has been upregulated ([307]; [308]). To implement a silencing delay, we assumed that Xist RNA must transit through a number of intermediate states x', x'', x'''... with a rate of 1 h^{-1} before reaching the silencing competent states x_1 and x_2 where it can induces changes at the cXR or tXA promoter that reduce the production rate of the respective regulator (Fig 9.1). To this end, we introduced two additional parameters that determine the number of intermediate states through which the Xist RNA must transit before being able to silence cXR ($sil_{cXR} = p_7$) and tXA ($sil_{tXA} = p_8$), respectively.

FIGURE 9.1: Scheme of the silencing delay implementation. sil_{cXR} or sil_{tXA} determine the number of silencing steps that exist and correspond to the mean silencing delay.

9.2.2 Simulation of monoallelic Xist upregulation

Parameter Values

We repeated the ODE simulations in section 9.1 for 300,000 randomly sampled parameter combinations to obtain more parameter sets that fulfilled the post-XCI requirements. Then we combined each parameter set that could stabilize the XaXi state in females and the Xa state in males, while destabilizing the XiXi state in females and the Xi state in males, with 10 random combinations of silencing delays and scaling factors. The silencing delays p_7 (sil_{cXR}) and p_8 (sil_{tXA}) were drawn from a uniform distribution between 1 and 20, while the scaling factors were drawn from a logarithmic distribution between 50 and 500. Table 2.5 in section 2.2 summarizes the parameters of the resulting model.

Initial conditions and simulation details

For each parameter set 100 cells were simulated for 100 h starting from an XaXa state with $x_1 = x_2 = 0$, $cXR_1 = cXR_2 = p_{22}$, and $tXA_1 = tXA_2 = p_{23}$. The simulations were performed in Julia using the Gillespie algorithm and run on a computing cluster [135, 136].

Simulation analysis

For each hour of the simulation each chromosome was classified as Xist positive (Xist+) if the mean Xist level exceeded 20% of the $Xist^{high}$ state from the ODE simulations ($0.2 \cdot Xist^{high} \cdot p_{21}$), and otherwise as Xist negative (Xist-). From this, the fraction of cells displaying mono- or biallelic Xist expression was calculated.

9.2.3 Allelic and cellular steady state analysis

To understand how the two regulators impact monoallelic Xist upregulation, we analyzed the steady states of the system, both locally at the allele level and globally at the cell level, in the presence of both, the cXR-, and the tXA- mediated feedbacks, and upon removal of each of them. To analyze the allelic steady states we performed an ODE simulation as in section 9.1 starting from different initial conditions of Xist and cXR (Xist = 0, 0.1, 0.2, ...1; cXR = 1- Xist), and for different tXA doses, which remained constant during the simulation.

The plots in Fig 2.4 show the allelic Xist steady states reached after 100 h of simulation. The tXA level is scaled to the level that would correspond to 0, 1 or 2 active X chromosomes (XiXi, XaXi, XaXa), as calculated from the $Xist^{high}$ and $Xist^{low}$ steady states in the original ODE simulation of the full model (see section 9.1):

$$tXA_{Total} = tXA_{Xa} + tXA_{Xi}$$

$$\text{with } tXA_{Xa} = 1 - \frac{(Xist^{low})^{p3}}{(Xist^{low})^{p3} + p4^{p3}} \text{ and } tXA_{Xi} = 1 - \frac{(Xist^{high})^{p3}}{(Xist^{high})^{p3} + p4^{p3}}$$

To perturb the negative feedback, only a single constant tXA concentration was simulated corresponding to the single dose present in the XaXi state. This reflects that the tXA dose becomes independent of the Xist expression state on the two chromosomes, if tXA escapes silencing.

To block the cXR-mediated arm of the local double negative feedback, we set the threshold level that describes how much cXR is necessary to reduce Xist by half to 1000, so that there is effectively no cXR repression of Xist.

To identify the global steady states of the entire cell with two X chromosomes for the full model and the two feedback-blocked models, we simulated all combinations of initial values for the two Xist alleles (Xist1 = 0, 0.1, 0.2, ..., 1 and Xist2 = 0, 0.1, 0.2, ..., 1). The initial values of cXR_1, cXR_2, tXA_1 and tXA_2 were set to the steady state values that result from the respective $Xist_1$ and $Xist_2$ values:

$$\frac{dcXR_i}{dt} = (1 - \frac{Xist_i^{p5}}{X_i^{p5} + p6^{p5}}) - cXR_i$$

$$\frac{dtXA_i}{dt} = (1 - \frac{Xist_i^{p3}}{X_i^{p3} + p4^{p3}}) - tXA_i$$

For the blocked negative feedback the total tXA concentration was kept constant as described above. The Xist steady states reached after 100 h of simulation are plotted in Fig 2.4 (bottom) in the section 2.2. The global steady states that could only be reached from a symmetric initial condition ($Xist_1 = Xist_2$) are indicated as unstable.

9.2.4 Reproducing experimental measurements of Xist upregulation

To investigate whether the model could also reproduce quantitative Xist expression data we compared the simulations to RNA FISH data from differentiating female mESCs and to in vivo

data in mouse and rabbit ([99, 142]). To temporally align simulation and experiment, different values for the time point when random XCI initiates were tested (offset), within a 24 h time window before the first time point when Xist clouds were observed. To identify the parameter set - offset combination that explains the data best we used a maximum likelihood estimate (MLE) approach. Both data and simulations were modeled as a multinomial distribution (mnpdf). The log Likelihood was then calculated as follows:

$$MLE = \sum_t log \sum_i mnpdf(D_t|p_i) \cdot mnpdf(S_t|p_i)$$

with i = 0, 0.01, ..., 1. t represents the measured timepoints and D_t and S_t are vectors containing the number of cells with no, monoallelic and biallelic Xist expression in the data and the simulation, respectively. The parameter set - offset combinations that maximized the likelihood function were selected. A summary of the data sets used and the window in which the offset was tested as well as the selected offset are given in table 9.5. The selected parameter sets that were used in Fig 2.9f are given in table 9.6.

Table 9.5: Summary of the data sets that were compared to the model simulation

Data set	**Time points**	**Offset window tested**	**Offset selected**
mESCs	0, 1, 2, 3, 4 days	0 ... 24 h	9 h
Mouse embryos	E5.0, E5.5, E6.5, E7.5	96 ... 120 h	106 h (∼E4.4)
Rabbit embryos	Morula (67), 96, 120 h.p.c	43 ... 67 h	57 h

Table 9.6: Parameter sets that best explain the experimental data

Data set	p_3	p_4	p_5	p_6	p_7	p_8	p_{11}	p_{12}	p_{13}	p_{14}	p_{21}	p_{22}	p_{23}
mESC	1.5	0.023	3.7	0.026	1	16	2.5	0.64	1.9	0.10	75	167	151
mouse	3.4	0.017	2.2	0.019	1	5	2.7	1.03	2.6	0.20	347	76	79
rabbit	2.2	0.048	3.0	0.047	6	12	1.7	1.4	2.3	0.35	451	207	220

9.2.5 Simulating aneuploid and polyploid cells

Simulation of Xist upregulation in cells with one, three or four X chromosomes was conducted as in section 9.2.2, except that the number of X chromosomes in each cell was modified accordingly, each contributing a single tXA dose. We used the 100 parameter sets that could best explain mouse in vivo data (see section 9.2.4) to simulate 100 cells for each aneuploid genotype. Again, each chromosome was classified as Xist positive (Xist+) if the mean Xist level exceeded 20% of the $Xist^{high}$ state from the ODE simulations ($0.2 \cdot Xist^{high} \cdot p_{21}$), and otherwise as Xist negative (Xist-) for each hour of the simulation. We then classified for each cell the number of inactive X chromosomes (no Xi, Xi, XiXi, XiXiXi or XiXiXiXi) based on how many Xist+ alleles were present at the end of the simulation (Fig 2.6b).

To simulate polyploid cells, we assumed that either tXA is diluted due to an increased nuclear volume (Fig 2.6a), or that tXA is repressed by autosomal factors (Fig 2.6c). To simulate tXA dilution we adapted the effect of tXA on Xist as follows:

$$\frac{dx}{dt} = p_{21} \cdot f(cXR) \cdot \frac{(\frac{1}{p_{33}}(\sum_{i=1}^{n} tXA_i))^{p_{11}}}{(\frac{1}{p_{33}}(\sum_{i=1}^{n} tXA_i))^{p_{11}} + (p_{23} \cdot p_{12})^{p_{11}}} - x$$

n specifies the number of X chromosomes and p_{33} specifies the autosomal ploidy ($p_{33} = 3$ in triploid and $p_{33} = 4$ in tetraploid cells). The 3 Xs in triploid and the 4 Xs in tetraploid cells therefore produce the same tXA concentration as the 2 Xs in diploid cells.

To simulate tXA repression by an autosomal factor, we modified tXA production as follows:

$$\frac{dtXA}{dt} = p_{23} \cdot \frac{2}{p_{33}}(1 - \frac{x^{p_3}}{x^{p_3} + (p_{21} \cdot p_4)^{p_3}}) - tXA$$

9.2.6 Simulating Xist upregulation in human embryos

The equations for the simulation of (1) cXR upregulation and (2) cXR dampening scenarios can be found in Table 2.6 in section 2.4. We simulated both scenarios with the 100 parameter sets that could best explain mouse in vivo data (see section 9.2.4).

9.3 Stochastic antisense simulations

We developed a mathematical model simulating RNAP traffic originating from two convergent genes. This model was used for the simulation of the *Xist/Tsix* locus and for the simulation of a general antisense locus (Fig 2.14c and Fig 5.1). In the model, two convergent promoters drive expression from sense and antisense strand, respectively. The model describes transcription initiation, transcription elongation, and RNA degradation of the two genes. It tracks binding and movement of RNAPs along the DNA. Both promoters can exist in a transcriptionally inactive OFF state, and an active ON state, where transcription can be initiated with constant rates. To describe transcription elongation, the gene bodies were devided into segments of 100 nt and RNAPs move along the gene body with a constant rate k_{elong}. RNAPs that reach the end of their gene produce one RNA molecule. Degradation of the RNA obeys first-order reaction kinetics with constant degradation rates k_{deg}. The two antisense genes mutually interfere with their transcription in the following ways:

- **RNAP collisions:** Two opposing RNAPs that occupy the same DNA segment collide. Upon a collision exactly one randomly chosen RNAP dislodges from the DNA.
- **Occlusion:** Binding of new RNAPs to a promoter is prevented if the promoter is occupied by an RNAP that initiated from the convergent promoter, but also if the previously recruited RNAP has not left the promoter yet (self-occlusion).
- **Promoter repression:** If a gene extends over its convergent promoter, passing RNAPs can turn off the convergent promoter. The OFF state is reverted back to the ON state with a constant rate.

Additional mechanisms of interference and other details specific to either the Xist/Tsix model or the general antisense model are described in section 9.3.1 and 9.3.2, respectively.

Simulations

All simulations were conducted in MATLAB. The model was written in C++ and compiled into a

MEX file that was called from the main MATLAB function. For parameter scanning a compiled Matlab script was executed in parallel on a computing cluster. In the simulation, transcription elongation occurs at fixed time intervals of 2.5 seconds based on measurements of polymerase speed (elongation of one 100 bp interval at k_{elong}=40 bp/sec [149]). Between elongation steps, all other reactions are simulated using the stochastic Gillespie algorithm [136]. Per parameter set, 100 realizations were performed.

9.3.1 Xist/Tsix model

The Xist/Tsix model is summarized in Fig 2.14c. Table 2.10 summarizes the parameters of the Xist/Tsix model. In addition to the mechanisms of transcriptional interference described above, Xist RNA silences the *Tsix* and *tXA* promoters. This was implemented in the following way: If Xist RNA is present above a threshold level of 10 RNA molecules it induces a transition of the *Tsix* and *tXA* promoters to the OFF state. To account for the delay with which Xist RNA silenced X-linked genes, we implemented a silencing delay described by the parameters sil_{tXA} or sil_{Tsix}. To this end, each chromosome passes stochastically through a number of intermediate states S_1, S_2...S_n once Xist expression from that chromosome has exceeded the threshold of 10 molecules and gene silencing occurs once the final silencing state S_n has been reached. If the level of Xist RNA molecules drops below the threshold before S_n has been reached, the chromosome immediately passes back to the unsilenced state S_0. The transitions through the intermediate states occur with rate 1 h^{-1} such that the number of intermediate states given by the model parameters sil_{tXA} or sil_{Tsix} is equal to the mean silencing delay. Silencing is assumed to be reversed, if the Xist level drops below the threshold of 10 molecules. Reactivation of Tsix and tXA will then occur with a single stochastic reaction with the rates $k_{react-T}$ and $k_{react-tXA}$ respectively (Fig 2.17b).

To account for regulation of Xist by tXA, the tXA dosage was assumed to modulate the effective Xist initiation rate k_X^{eff} as follows: $k_X^{eff} = q_{tXA} \cdot k_X$, where k_X is the Xist initiation rate in the presence of a single tXA dose and $q_{tXA} = 0, 1, 2$ depending on whether no, one or two tXA loci are active. The tXA concentration was modeled as a step function with the value 1 if the respective tXA allele is active and the value 0 if the respective tXA allele has been silenced by Xist RNA. The kinetics of RNA and protein decay of tXA were not accounted for explicitly but were instead assumed to modulate the tXA silencing kinetics (see below). To reproduce coupling of XCI to development, k_X^{eff} must also be influenced by the differentiation timing, representing the action of stem cell-specific factors that prevent Xist upregulation in undifferentiated cells by repressing Xist. The differentiation dependency was formulated as a step function that changes its value at the point of induction of differentiation such that k_X^{eff} prior to differentiation was 10% of the k_X value after the onset of differentiation.

Before onset of differentiation: $k_X^{eff} = 0.1 \cdot q_{tXA} \cdot k_X$
After onset of differentiation: $k_X^{eff} = 1 \cdot q_{tXA} \cdot k_X$

Reduced models

To analyze which of the repressive mechanisms are strictly required for establishment and maintenance of the XaXi state, we systematically investigated reduced model structures with a combination of two or only a single repressive mechanism. These simplifications were implemented in the following way:

- In all models without *Xist* promoter repression, passing Tsix polymerases do not affect the *Xist* promoter state.
- In all models without *Tsix* promoter silencing, the Xist RNA does not affect the *Tsix* promoter state.
- In all models without polymerase collisions, Xist and Tsix polymerases were assumed to be able to bypass each other.

9.3.1.1 Simulating maintenance of the XaXi state

Parameter values

Table 2.10 in section 2.6 summarizes all parameters with their respective ranges. RNA degradation and elongation rates were set to fixed values based on previous experimental estimates. Silencing has already occurred in the XaXi state, so that silencing kinetics should not affect the outcome of the simulation. Silencing and reactivation rates were therefore set to constant values (1 h^{-1}). Similarly, tXA is present at a constant single dose in post-XCI cells with a single Xa and is therefore set to a constant value of 1. All other parameters were varied within realistic parameter ranges and systematically combined resulting in 8000 parameter sets in total.

Initial conditions

To investigate Xa and Xi stability, each parameter set was simulated with two asymmetric initial conditions, corresponding to the Xa and Xi transcription states. On the Xa, Tsix is transcribed while Xist is silent. On the Xi, Xist is transcribed while Tsix is transcriptionally silent. The polymerases on the transcribed genes were distributed randomly across the gene body with steady state polymerase occupancy:

$$N_{RNAP} = \frac{L_{gene}}{k_{elong}} \cdot k_{ini}$$

where L_{gene} denotes the gene length, k_{elong} the elongation rate, and k_{ini} the initiation rate. RNA levels of transcribed genes were set to their steady state value.

Table 9.7: Initial Conditions XaXi

	Xi	Xa
Xist RNA	$\frac{k_X}{k_{deg-X}}$	0
Tsix RNA	0	$\frac{k_T}{k_{deg-T}}$
Xist promoter	ON	OFF
Tsix promoter	OFF	ON
No. Xist RNAPs	$\frac{L_{Xist}}{k_{elong}} \cdot k_X$	0
No. Tsix RNAPs	0	$\frac{L_{Tsix}}{k_{elong}} \cdot k_T$

Simulation

For each parameter set 100 Xi/Xa pairs were simulated for 500 h to reach the steady state.

Simulation Analysis

A simulation was classified as stably maintaining the XaXi state if Xist was on average present with >10 molecules at the Xi and with <10 molecules at the Xa during the last 50 h of the simulation. Parameter sets, where >99% of Xa/Xi pairs stably maintained the XaXi state were classified as monoallelic.

Based on systematic variation of k_X two threshold were defined: A lower "activation" threshold, below which >99% of Xa/Xi pairs were stable, and an upper threshold, above which <1% of Xa/Xi pairs were stable (2.18d-g, red and grey).

Reduced Models

The reduced models were simulated and analyzed as described above with the parameters values for k_X, k_T and $k_{rev-rep}$ (where applicable) given in Table 2.10. Variation of the half-life of the repressed *Xist* promoter state in the reduced models, allowed us to compare the effect of its stability in full and the reduced model without polymerase collisions (see Fig 2.16c). Parameters as given in Table 2.10 were systematically combined, resulting in 8000 parameter sets for all models with 3 variable parameters, and 400 sets for all models with 2 variable parameters. A summary of the simulations of all reduced models is given in Table 9.8.

Table 9.8: Reduced models XaXi maintenance

Model	Parameters	# parameter sets	XaXi stable [%]
Silencing of Tsix, Xist repression, RNAP collisions	k_X, k_T, $k_{rev-rep}$	8000	50%
Xist repression, RNAP collisions	k_X, k_T, $k_{rev-rep}$	8000	0%
Silencing of Tsix, RNAP collisions	k_X, k_T	400	45%
Silencing of Tsix, Xist repression	k_X, k_T, $k_{rev-rep}$	8000	26.4%
Silencing of Tsix	k_X, k_T	400	0%
Xist repression	k_X, k_T, $k_{rev-rep}$	8000	0%
RNAP collisions	k_X, k_T	400	0%

Partial Tsix silencing

To assess, whether partial silencing of Tsix is sufficient to generate bistability we varied the strength of Xist-RNA mediated silencing: Instead of inducing a transition of the *Tsix* promoter to the OFF-state, Xist RNA is assumed to reduce the Tsix initiation rate by a certain fraction given by the model parameter $sil_{partial}$ ($sil_{partial} = 1$ complete silencing). The model was simulated and analyzed as described above with the parameters values given in Tables 2.10 and 9.9.

Table 9.9: Parameter for simulation of partial Tsix silencing

Description	Parameter	Parameter Values
Fraction of k_T reduction	$sil_{partial}$	0, 0.1, 0.2, ..., 1

Reduced Overlap

To simulate alternative locus structures, the *Tsix* gene was assumed to be truncated at different positions upstream of the *Xist* promoter. For each of these scenarios a systematic parameter analysis was performed as described in the previous section and the activation thresholds were determined (Fig. 2.16d). For each locus architecture the k_X/k_T ratio at the activation threshold is given in the Table 9.10.

TABLE 9.10: k_X/k_T at the activation threshold

Overlap	For bistability: $k_X/k_T \leq$
23 kb	0.91
16 kb	0.87
8 kb	0.77
4 kb	0.69
2 kb	0.56
1 kb	0.22

9.3.1.2 Simulating XaXa → XaXi transition

Parameter values

Xist upregulation was now simulated for all parameter sets that could stably maintain the XaXi state in the maintenance simulation in 9.3.1.1 (4001 sets). Each parameter set was combined with 500 combinations of randomly sampled values for sil_{tXA}, sil_{Tsix}, $k_{react-tXA}$ and $k_{react-T}$ (see Table 2.11).

Initial Conditions

Both chromosomes were initiated from the Xa state (see Table 9.7) in undifferentiated cells with double tXA dosage present (q_{tXA}=2).

Simulation

100 cells were simulated for each parameter set. To reach the steady state prior to differentiation, the cells were simulated for 10 h in an undifferentiated state, then 100 hours of differentiation were simulated.

Simulation analysis

Each cell was classified as monoallelic, if during the last 20 h of the simulation >10 molecules of Xist RNA were present on average at one chromosome (Xi) and <10 molecules on the other (Xa).

Reduced Models

To analyze which repressive mechanisms were required for the establishment of the XaXi state, Xist upregulation was also simulated with the reduced models using the parameter sets that could maintain the XaXi state, each combined with 500 combinations of randomly sampled values for sil_{tXA}, sil_{Tsix}, $k_{react-tXA}$ and $k_{react-T}$. A summary of the simulations of the full model and of all reduced models is given in Table 9.11. The reduced model without Tsix-mediated repression of the *Xist* promoter was termed the "antisense model" and used for all subsequent simulations.

Table 9.11: Reduced models XaXa → XaXi

Model	# parameter sets	Monoallelic sets
Silencing of Tsix, Xist repression, RNAP collisions	2000500	1.17% MA in >99% of cells
Silencing of Tsix, RNAP collisions	90000	1.53% MA in >99% of cells
Silencing of Tsix, Xist repression	1057500	0% MA in >99% of cells

Simulation of reduced overlap

To investigate whether the antisense model could in principle still ensure robust monoallelic Xist upregulation if the overlap between *Xist* and *Tsix* was reduced, we repeated the analysis for the human *XIST/TSIX* locus architecture with 8 kb overlap (Fig 2.19). For all parameters that could maintain the XaXi state, Xist upregulation was simulated as described in section 9.3.1.2. A summary of the simulations is given in table 9.12 and in Fig 2.19.

Table 9.12: Variation of Locus architecture

Overlap	parameters varied	XaXi → XaXi		XaXa → XaXi	
		No. sets tested	XaXi stable	No. sets tested	XaXi established
23 kb	k_X, k_T	400	45%	90,000	1.53%
8 kb	k_X, k_T	400	38%	76,000	0.85%

9.3.1.3 Simulating different genotypes

Identification of parameter sets compatible with experimental data

To simulate experimental data we selected a subset of parameter sets from the simulation in 9.3.1.2 that robustly led to monoallelic Xist upregulation (>99% cells) and were in agreement with experimental observations, as described in section 2.6.4. Since only 34 parameter sets with unique k_X, k_T, sil_{tXA} and sil_{Tsix} combinations fulfilled these criteria, we performed another simulation to identify more parameter sets that could potentially simulate experimental data. To this end, the simulation in 9.3.1.2 was repeated with additional, randomly sampled values for k_X and k_T. The parameter ranges for k_X were set such that the steady state expression level of Xist (k_X/k_{X-deg}) was between 200 and 600 molecules. Since the simulation in 9.3.1.2 revealed that monoallelic Xist upregulation requires a k_X-to-k_T ratio between 0.4 and 0.8 (Fig 2.18c), k_T was sampled within this range. A total of 500,000 parameter sets were simulated. All parameters were sampled randomly within the ranges given in the following table.

Table 9.13: Parameter Values Mutant simulation

Description	Parameter	Parameter Values
Xist initiation rate [h^{-1}]	k_X	34 - 104 (log distributed)
Tsix initiation rate [h^{-1}]	k_T	$k_X/0.8$ - $k_X/0.4$ (lin distributed)
Silencing delay of tXA [h]	sil_{tXA}	0 - 48 (log distributed)
Silencing delay of Tsix [h]	sil_{Tsix}	0 - 48 (log distributed)
Reactivation rate of tXA [h^{-1}]	$k_{react-tXA}$	0.1 - 100 (log distributed)
Reactivation rate of Tsix [h^{-1}]	$k_{react-Tsix}$	0.1 - 100 (log distributed)

From these simulations 100 sets fulfilling the above requirements were randomly selected and

were used in the following simulations.

Simulation of Xist and Tsix mutants
Xist and Tsix mutations were simulated as described in section 9.3.1.2 with adapted Xist and Tsix initiation rates as described in section 2.6.4. To estimate the half-time of Xist upregulation $T_{1/2}$ shown in Fig 2.25e-f we determined the earliest time point where 50% of simulated cells had upregulated Xist (>10 molecules).

Simulation of aneuploid and polyploid cells
To simulate cells that are mono-, tri- or tetrasomic for the X chromosome (Fig 2.27a), simulations were performed essentially as described in section 9.3.1.2 except that each cell contained one, three or four X chromosomes, each contributing a single tXA-dose. To simulate tetraploid cells (4n4X, Fig 2.27b) we assumed that their increased nuclear size would result in an effective dilution of tXA compared to diploid cells. Each of the four X chromosomes in a cell therefore produces only 0.5x tXA.

To identify the parameter sets that are predicted to inactivate exactly two of the four X chromosomes (XiXaXaXa unstable, blue in Fig 2.27b), we performed a maintenance simulation as described in section 9.3.1.1 for all mutant parameter sets (500h, 100 cells per set) with an tXA dosage of 0.5, 1, 1.5 and 2 to determine which of the sets were bistable with 1x tXA and monostable with a 1.5x tXA concentration.

9.3.2 General antisense model

In essence, transcription was modelled as described in section 9.3, with the following modifications, to dissect the effect of different regulatory events involved in transcription initiation and transcription interference:

- **Promoter-proximal pausing:** Transcription initiation was modelled as a two-step process, where RNAPs bind to the promoter region with a binding rate k_{bind}, forming a paused initiation complex. Pause release occurs with rate $k_{release}$ and either results in spontaneous termination of RNAPs with probability p_{term} or in the formation of an elongation complex with probability $1 - p_{term}$. Elongation complexes then start productive elongation. While the promoter is occupied by a paused RNAP, no new RNAPs can bind (self-occlusion).
- **Sitting duck interference (SDI):** While a promoter-bound RNAP is in the paused state, it interferes with elongating antisense RNAPs by sitting duck interference, which results in dislodgment of the paused RNAP.
- **Multiple promoter OFF states:** Antisense RNAPs that transcribe over the sense promoter segment, induce the transition of the promoter to an inactive state where no transcription can be initiated. To account for the refractory period in mammalian gene reactivation, we included multiple promoter OFF states [261–263]. Promoter transitions are modelled as an irreversible cycle of one active state (ON state) where transcription can initiate and N sequential inactive OFF states (promoter progression model, see [264, 265]). We assume that the mean time spent per state is the same for each inactive state, such

that the rates of transition to the next state are equal and determined as $k = \frac{N}{t_{OFF}}$, where N is the number of OFF states and t_{OFF} is the mean total silent period.

9.3.2.1 Simulating transcription state maintenance

Parameter values

Table 5.1 in section 5.1 summarizes the parameters of the general antisense model.

Initial Conditions

To understand whether TI alone could stably maintain alternative transcription states, each allele was simulated twice, starting from asymmetric initial conditions:

- Initial condition 1: Gene A transcribed at steady state in the absence of TI, gene B OFF
- Initial condition 2: Gene B transcribed at steady state in the absence of TI, gene A OFF

In the absence of transcriptional interference by the antisense gene, the steady state occupancy of RNAPs on the gene can be inferred as:

$$\sum RNAP = \frac{L_{Gene}}{k_{elong}} \cdot (\frac{1}{k_{bind}} + \frac{1}{k_{release}})^{-1} \cdot p_{term}$$

where $\frac{L_{Gene}}{k_{elong}}$ corresponds to the time that an elongating RNAP needs to transcribe the gene, and $(\frac{1}{k_{bind}} + \frac{1}{k_{release}})^{-1} \cdot p_{term}$ gives the net initiation rate, with $\frac{1}{k_{bind}} + \frac{1}{k_{release}}$ corresponding to the average time the promoter needs to produce one elongating RNAP.

The steady state RNA level can then be inferred as the ratio of net production and degradation rates:

$$\sum RNA = \frac{(\frac{1}{k_{bind}} + \frac{1}{k_{release}})^{-1} \cdot p_{term}}{k_{deg-RNA}}$$

Simulation

The simulations were conducted as described in section 9.3.1.1. For each parameter set the antisense locus was simulated from both initial conditions for 500h, in 100 realizations.

Simulation analysis

As a classifier for the stability of the initial transcription state the first-switching-time was extracted. The first-switching-time denotes the first time point at which the initially inactive gene dominated transcription within the overlap. It was defined as the first time point at which the ratio of the sum of RNAPs within the overlap, normalized to the relative promoter strength of the two genes, crossed 1:

$$\frac{\sum RNAP_A/k_{net-Ini-A}}{\sum RNAP_B/k_{net-Ini-B}}, \text{ with } k_{net-Ini} = (\frac{1}{k_{bind}} + \frac{1}{k_{release}})^{-1} \cdot p_{term}$$

The first-switching-time was then averaged over all simulated alleles of the same parameter set,

and the minimum over the two initial conditions was used to measure transcriptional stability. This measure was termed minimal first-switching-time (MFS).

A set was classified as stable if $MFS = min(\overline{fsw}_{Ini-1}, \overline{fsw}_{Ini-2}) > 200\text{h}$.

Reduced models

To analyze which TI mechanisms and transcription reactions were strictly required for the generation of transcriptional memory, we generated five reduced model versions, each either lacking one of the TI mechanisms or simplifying transcription initiation or promoter re-activation into single-step reactions.

- In the model without promoter repression, passing RNAPs do not affect the state of the opposing promoter.
- In the model without collisions, sense and antisense RNAPs can bypass one another
- In the model without SDI, both EC and SDI have the same probability to dislodge upon collision
- To simplify promoter reactivation into a single-step reaction, we set the number of promoter OFF states N to 1.
- To simplify transcription intiation into a single-step reaction, we set $k_{release} = 100,000\ \text{h}^{-1}$ and $p_{term} = 0$ such that every RNAP that binds the promoter is immediately released into productive elongation. The new lumped one-step initiation rate k_{ini} was sampled between [5,500] h^{-1}

Each simplified model was simulated with 15,000 randomly sampled parameter sets. Table 5.2 summarizes the reduced models.

Simulation of maximally simplified model

We resimulated the maximally simplified model with a larger number of parameter sets (50,000). Since it is unclear how frequently RNAPs collide in vivo, or induce a repressed promoter state, we included two additional parameters p_{PR} and p_{Coll} that modify the frequency of these events, respectively. Parameter values were randomly drawn from uniform or logarithmic distributions as indicated in Table 9.14.

Systematic parameter variation

100 stable sets (MFS>200h) were chosen at random, and in each of those one parameter was varied at a time, testing each 50 randomly chosen values between the original value and the upper and lower bound, respectively.

Table 9.14: Parameter Values of the simplified antisense model

Description	Parameter	Parameter Values
Initiation rate Gene A [h^{-1}]	k_{ini-A}	5 - 500 (log distributed)
Initiation rate Gene B [h^{-1}]	k_{ini-B}	5 - 500 (log distributed)
Average time of pA in OFF state [h]	t_{OFF-A}	1/60 - 10 (log distributed)
Average time of pB in OFF state [h]	t_{OFF-B}	1/60 - 10 (log distributed)
Probability for collision	p_{Coll}	0 - 1 (lin distributed)
Probability for promoter repression	p_{PR}	0 - 1 (lin distributed)
Overlap between A and B [kb]	L	5 - 50 (lin distributed)

Part V

Appendix

A Rules for monoallelic Xist upregulation in the Xist/Tsix model

Here, we set out to establish parameter rules that would allow us to predict whether a specific parameter set will generate monoallelic Xist upregulation. To this end, we defined parameter sets as monoallelic, if they could produce monoallelic Xist expression in >90% of simulated cells. As established in the section 2.6.2, a cell must cross the k_X-to-k_T activation threshold in the presence of a double tXA dose to trigger Xist upregulation. Moreover, the lower the k_X-to-k_T ratio the slower the switch-on of Xist (see Fig 2.21). The k_X-to-k_T ratio must thus be high enough to pass the activation threshold and to guarantee Xist upregulation within the relevant time frame (Fig 9.2, Group1 vs Group2; orange scenario in Fig 2.22). Among the sets that achieve Xist upregulation the relative timescales of Xist switch-on and tXA silencing (sil_{tXA}) determine whether Xist is upregulated monoallelically (group 1.1), or biallelically (group 1.2). If Xist is upregulated monoallelically, it is usually stably maintained (group1.1.1) unless tXA is silenced faster than Tsix (sil_{tXA}/sil_{Tsix} <1), a scenario that rapidly destabilizes Xist expression (see switch-off time Fig 2.21b). If now Tsix fails to be silenced before the Xist RNA is completely degraded, Xist expression cannot be stabilized (Fig 9.2, Group 1.1.2, red scenario in Fig 2.22). Whether this happens depends on the amount of Xist RNA and therefore correlates with k_X.

In the sets that upregulate Xist biallelically (group 1.2), the symmetry break between the two X chromosomes must occur through monoallelic silencing or reactivation of Tsix (group 1.2.5). We identified several scenarios in which this fails:
1) A symmetry break by monoallelic silencing is prevented if Tsix is silenced much faster than tXA ($sil_{tXA} - sil_{Tsix}$ >10, group 1.2.1 and 1.2.2). Symmetry break can then still occur by monoallelic reactivation unless tXA silencing or tXA reactivation are extremly slow. If tXA silencing is too slow cells cannot revert biallelic Xist expression and undergo a second attempt of monoallelic upregulation within the relevant timescales (Fig 9.2 group 1.2.1, green scenario Fig 2.22). If tXA reactivation is too slow, Tsix will always be reactivated on both alleles before tXA is reactivated, preventing a symmetry break (Fig 9.2 group 1.2.2, dark blue scneario in Fig 2.22).
2) Symmetry break also fails if tXA is silenced much faster than Tsix ($sil_{Tsix} - sil_{tXA} >$ 10), and all produced Xist RNA is degraded upon tXA silencing before Tsix can be silenced ($k_X/(sil_{Tsix} - sil_{tXA}) < 25$, Fig 9.2 group 1.2.3, light blue scenario in Fig 2.22).
3) In cells that have silenced both tXA and Tsix copies, symmetry break will also fail if tXA is reactivated faster than Tsix: In the absence of *cis* repression of Xist by Tsix, Xist is immediately upregulated from both alleles, preventing symmetry break during reactivation (Fig 9.2 group 1.2.4, purple scenario in Fig 2.22). This is the case if $k_{react-tXA}/k_{react-Tsix} > 5$ or if both reactivation rates are very slow so that the order if reactivation is variable ($k_{react-Tsix} < 10$).

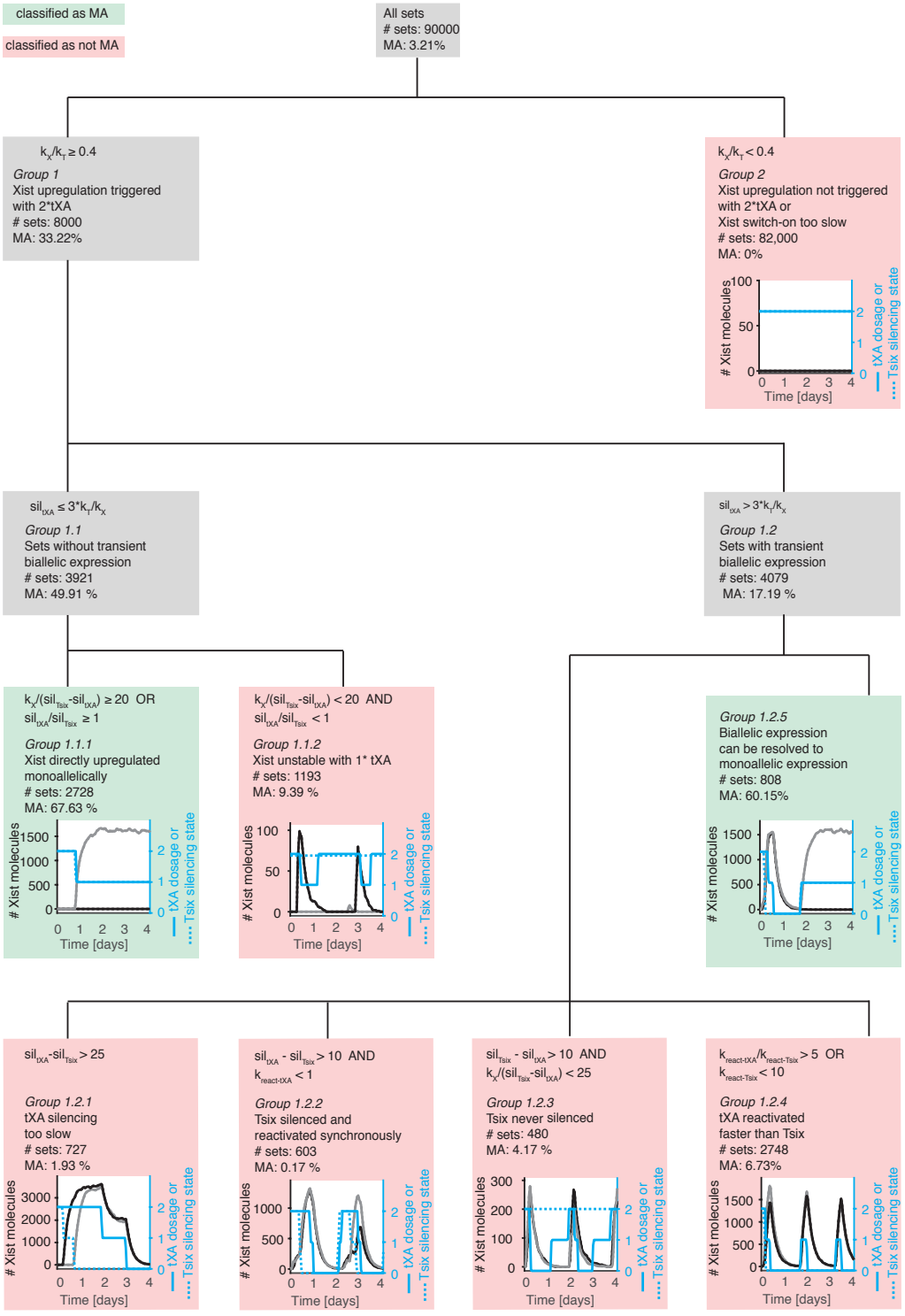

FIGURE 9.2: Parameter rules that classify sets into monoallelic or non-monoallelic. Sets were defined as monoallelic if >90% of cells were monoallelic (averaged over the last 20h of the simulation). Red shading: groups of sets that were classified as non-monoallelic, green shading: groups of cells that were classified as monoallelic. For each lowest level node a representative example simulation is shown. The number of sets and the percentage of monoallelic sets in each group is indicated.

B Transcriptional interference at the *Xist/Tsix* locus - replicates

Figure 9.3 depicts the three replicates of the experimental measurement of TI between Xist and Tsix described in section 2.6.3.

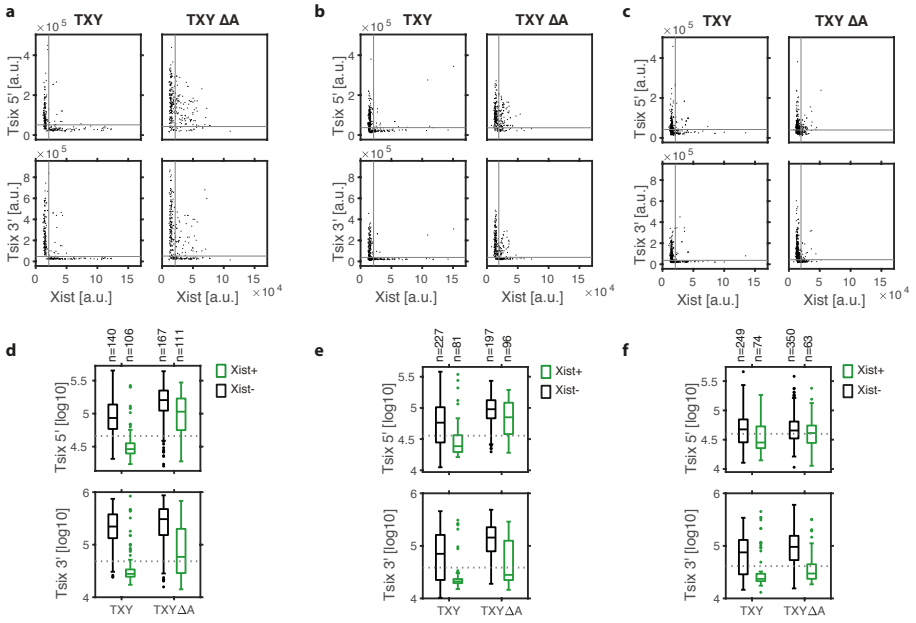

FIGURE 9.3: Three replicates of the TI experiment. TXY and TXYΔA ESCs were treated with doxycycline for 24 hours and nascent transcription of Xist and Tsix (5' and 3') was assessed by RNA FISH. (a-c) Quantification of 3 biological replicates, where each dot represents the measured signal intensities of a single allele. Grey lines indicate the detection threshold estimated from negative control regions. (d-f) Box plots of Tsix signal intensity at Xist+ (green) and Xist- alleles (black) in the two cell lines as indicated for the data shown in (a-c); dotted lines indicate the detection threshold (center line, median; box limits, upper and lower quartiles; whiskers, most extreme data points not considered outliers; points, outliers).

C Construction of a synthetic antisense locus

In this section, I describe the steps that we have taken towards the construction of a synthetic antisense locus that will be used to test the predictions of our antisense model experimentally.

As described in section 6.6, we designed the experimental assay in analogy to our theoretical analysis to probe for transcriptional memory. To this end, we needed a means of exposing the antisense locus to different initial conditions, and of reading out the locus' transcription state to analyze whether and for how long it remembers its initial transcription state (Fig 6.4).

We thus require two independent triggers that allow us to switch the state of the system by inducing either sense or antisense transcription. The host mESC line already contains the doxycycline inducible activation rtTA system (reverse tetracycline-controlled transactivator). In addition, we stably integrated a second system that relies on the transcription activation by the yeast TF Gal4 fused to a VP16 transactivator domain, and allows control over its nuclear concentration. To this end, we lentivirally integrated a construct expressing Gal4-VP16 fused to the mutant estrogen ligand-binding domain ERT2. ERT2 ensures the cytoplasmic localization of the "synthetic" TF (transcription activator), preventing constitutive activation of target sequences. In the presence of 4-hydroxy-tamoxifen (4-OHT) nuclear translocation of the fusion protein is induced [309]. We tested both systems by reading out expression levels of a lentivirally integrated GFP transgene under an inducible promoter with TetO or UAS motifs, and found that they are both capable of inducing strong expression (Fig 9.4c).

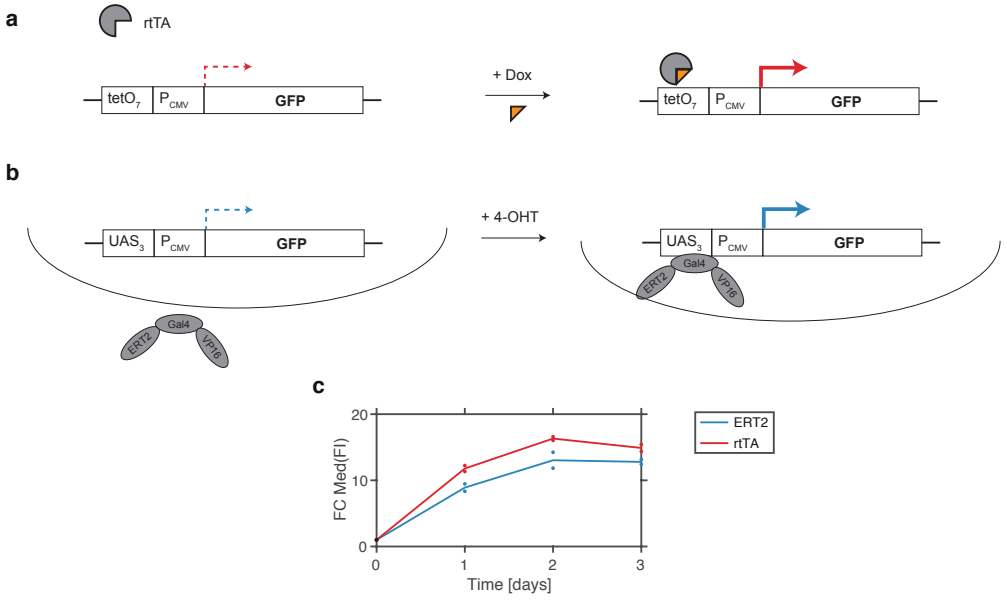

FIGURE 9.4: Testing two independent systems to induce transcription of sense and antisense strand using a GFP reporter under control of the different inducible promoters. a) Tetracycline inducible expression system. Upon treatment with doxycycline rTetR fused to the VP16 transactivator domain can bind tetO sequences within the inducible promoter and activate transcription. b) ERT2-Gal4-VP16. ERT2 allows control over the nuclear concentration of the Gal4-VP16 transcription activator. Upon 4-OHT induction the fusion protein translocates into the nucleus and is able to bind to UAS motifs within the inducible promoter to activate transcription. c) Time course of GFP induction over 3 days (4-OHT: 1μM, Dox: 2μg/ml). GFP expression was read out using flow cytometry. The fold change in median fluorescence intensity in the positive population is shown.

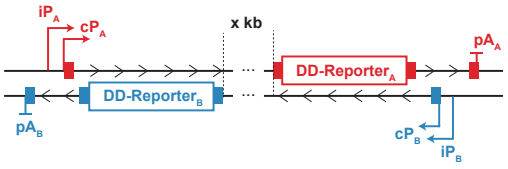

FIGURE 9.5: Design of the synthetic antisense construct. Both genes have a basal and an inducible promoter to allow control over the locus' transcription state. Transcription of both genes extends over the antisense promoter region. Insertion of splice sites allows modular extension of the inter-promoter distance, while guaranteeing translation. Each strand transcribes a fluorescent reporter that is coupled to a destabilization domain (ecDHFR) to allow direct monitoring of transcription.

To read out the transcriptional state of the antisense locus, we need to monitor transcription, ideally in live cells. Fluorescent reporter proteins can be used but they only monitor transcription indirectly, as they first have to be translated. High stability of the fluorescent reporter can mask fast fluctuating transcriptional changes. Therefore unstable reporter proteins are required that are still bright enough to be detected, if expressed from a single copy integration and under a potentially weak promoter. To achieve this we fused bright fluorescent proteins mAzami and mScarlet N-terminally to the FK506-binding protein (FKBP) L106P destabilization domain whose stability can be tuned with the small molecule ligand Shield-1 [310]. This destabilization domain is a mutant of the human FKBP12 protein that is unstable and confers this instability to its fusion proteins. To allow for the design of antisense constructs with a long inter-promoter distance, while still guaranteeing for efficient translation of the mRNA transcript, we decided to include intronic sequences, that will also allow us to monitor nascent transcription (e.g. by intronic RNA FISH). Specifically, we wanted the respective convergent reporter and promoters to be excised from the mature transcripts. To achieve this, we embedded the antisense reporter into the first intron of the human TATA Box binding protein (TBP) on the sense strand, and the sense reporter into the first intron of the alanyl-tRNA sythetase (AARS) gene on the antisense strand. Further, we inserted the exon2-exon3 splice junction of the respective genes downstream of the reporter, flanking the promoter of the convergent strand, to excise this region from the mature transcripts (Fig 9.5).

In order to test the conditions for transcriptional memory, multiple antisense constructs that differ in their locus architecture and promoter strength have to be tested in a comparable manner. This requires stable integration of the alternative constructs into the same target locus. As the target site for the integration of the antisense constructs we chose a “neutral” region on the murine chromosome 1, that does not contain active genes and is devoid of repressive or activating chromatin marks, but that does support transcription ([311], Luca Giorgetti personal communication, precise insertion point: chr1:73,641,436, mm10). To allow for straightforward site-specific single copy integration of constructs into this target locus, we have set up a BxB1-dependent landing pad system that is based on a design by Fonseca and colleagues [299]. To this end, we stably inserted a construct containing the *attP* recognition site between the *pCAG* promoter and a hygromycin resistance cassette into the target locus using CRISPR/Cas9 (Fig 9.6a). To now integrate constructs into this target locus via site-specific recombination they are cloned into a transfer vector, which also contains a promoter-less cassette encoding a blasticidin resistance together with a TagBFP downstream of the *attB* BxB1 site (Fig 9.6a). This allows for straightforward screening of the cells with correct integration of the constructs into the target site, as these cells will switch from being hygromycin-resistant to being blasticidin-resistant and in addition will gain TagBFP fluorescence (Fig 9.6b). We tested the landing pad by integrating a control construct containing a GFP transcription unit. Upon BxB1-mediated integration into the target locus and 10 days of blasticidin selection, the cells gain TagBFP and GFP fluorescence

(Fig 9.6c). However, a small fraction of cells remains GFP negative, possibly owing to incomplete integration events or epigenetic silencing in a subpopulation.

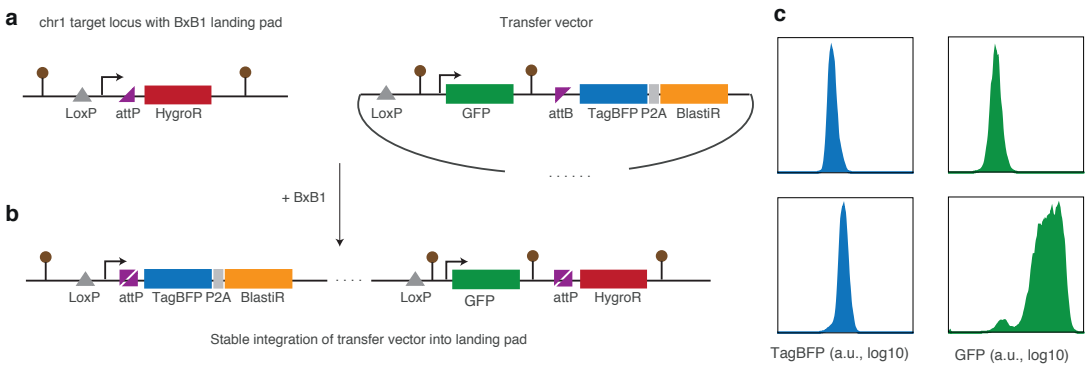

FIGURE 9.6: Generation and testing of a BxB1 dependent landing pad on chr1 in mESCs based on design from [299]. a) Scheme of the landing pad stably integrated in the chr1 target locus in mESCs (left) and of a positive control transfer vector containing a GFP transcription unit. b) Scheme of the locus after BxB1-mediated integration of the transfer vector. LoxP, LoxP site; attP, BxB1 phage attachment site; attP, BxB1 bacterial attachment site; HygroR, hygromycin resistance gene; BlastiR, blasticidin resistance gene; brown circles, chromatin insulators. c) GFP and TagBFP expression in the pool of parental cells containing the empty landing pad (top) and after BxB1-mediated integration of the transfer vector and 10 days of blasticidin selection (bottom).

We have thus set initial steps towards the construction of a synthetic antisense pair to probe for transcriptional memory experimentally.

Bibliography

1. Lyon, M. F. Gene Action in the X-chromosome of the Mouse (Mus musculus L.) *Nature* **190,** 372–373 (1961).
2. Gilfillan, G. D., Dahlsveen, I. K. & Becker, P. B. Lifting a chromosome: dosage compensation in Drosophila melanogaster. *FEBS Letters* **567,** 8–14 (2004).
3. Kuroda, M. I., Hilfiker, A. & Lucchesi, J. C. Dosage compensation in Drosophila - a model for the coordinate regulation of transcription. *Genetics* **204,** 435–450 (2016).
4. Albritton, S. E. & Ercan, S. Caenorhabditis elegans Dosage Compensation: Insights into Condensin-Mediated Gene Regulation. *Trends in Genetics* **34,** 41–53 (2018).
5. Ercan, S. Mechanisms of X Chromosome Dosage Compensation. *Journal of Genomics* **3,** 1 (2015).
6. Galupa, R. & Heard, E. X-Chromosome Inactivation: A Crossroads Between Chromosome Architecture and Gene Regulation. *Annual Review of Genetics* **52,** 535–566 (2018).
7. Rastan, S. Non-random X-chromosome inactivation in mouse X-autosome translocation embryos - location of the inactivation centre. *Development* **22,** 1–22 (1983).
8. Augui, S., Nora, E. P. & Heard, E. Regulation of X-chromosome inactivation by the X-inactivation centre. *Nature Reviews Genetics* **12,** 429–442 (2011).
9. Brown, S. D. M. XIST and the Mapping of the X Chromosome Inactivation Centre. *BioEssays* **13,** 607–612 (1991).
10. Nesterova, T. B., Slobodyanyuk, S. Y., Elisaphenko, E. A., Shevchenko, A. I., Johnston, C., Pavlova, M. E., Rogozin, I. B., Kolesnikov, N. N., Brockdorff, N. & Zakian, S. M. Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Research* **11,** 833–849 (2001).
11. Yen, Z. C., Meyer, I. M., Karalic, S. & Brown, C. J. A cross-species comparison of X-chromosome inactivation in Eutheria. *Genomics* **90,** 453–463 (2007).
12. Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nature Genetics* **30,** 167–74 (2002).
13. Loda, A. & Heard, E. Xist RNA in action: Past, present, and future. *PLoS Genetics* **15,** e1008333 (2019).
14. Giorgetti, L., Lajoie, B. R., Carter, A. C., Attia, M., Zhan, Y., Xu, J., Chen, C. J., Kaplan, N., Chang, H. Y., Heard, E. & Dekker, J. Structural organization of the inactive X chromosome in the mouse. *Nature* **535,** 575–579 (2016).

15. Minajigi, A., Froberg, J. E., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W. & Lee, J. T. A comprehensive Xist interactome reveals cohesin repulsion and an RNA- directed chromosome conformation. *Science* **349,** 282–293 (2015).

16. Darrow, E. M., Huntley, M. H., Dudchenko, O., Stamenova, E. K., Durand, N. C., Sun, Z., Huang, S. C., Sanborn, A. L., Machol, I., Shamim, M., Seberg, A. P., Lander, E. S., Chadwick, B. P. & Aiden, E. L. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proceedings of the National Academy of Sciences* **113,** E4504–E4512 (2016).

17. Deng, X., Ma, W., Ramani, V., Hill, A., Yang, F., Ay, F., Berletch, J. B., Blau, C. A., Shendure, J., Duan, Z., Noble, W. S. & Disteche, C. M. Bipartite structure of the inactive mouse X chromosome. *Genome Biology* **16,** 1–21 (2015).

18. Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S. & Aiden, E. L. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159,** 1665–1680 (2014).

19. Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N. L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J. & Heard, E. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–385 (2012).

20. Lee, J. T. & Lu, N. Targeted Mutagenesis of Tsix Leads to Nonrandom X Inactivation. *Cell* **99,** 47–57 (1999).

21. Stavropoulos, N., Lu, N. & Lee, J. T. A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice. *Proceedings of the National Academy of Sciences* **98,** 10232–7 (2001).

22. Luikenhuis, S., Wutz, A. & Jaenisch, R. Antisense transcription through the Xist locus mediates Tsix function in embryonic stem cells. *Molecular and Cellular Biology* **21,** 8512–8520 (2001).

23. Nesterova, T. B., Johnston, C. M., Appanah, R., Newall, A. E. T., Godwin, J., Alexiou, M. & Brockdorff, N. Skewing X chromosome choice by modulating sense transcription across the Xist locus. *Genes & Development* **17,** 2177–90 (2003).

24. Lee, J. T. Regulation of X-chromosome counting by Tsix and Xite sequences. *Science* **309,** 768–771 (2005).

25. Sado, T., Hoki, Y. & Sasaki, H. Tsix silences Xist through modification of chromatin structure. *Developmental Cell* **9,** 159–65 (2005).

26. Navarro, P., Pichard, S., Ciaudo, C., Avner, P. & Rougeulle, C. Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: Implications for X-chromosome inactivation. *Genes & Development* **19,** 1474–1484 (2005).

27. Navarro, P., Page, D. R., Avner, P. & Rougeulle, C. Tsix-mediated epigenetic switch of a CTCF-flanked region of the Xist promoter determines the Xist transcription program. *Genes & Development* **20,** 2787–2792 (2006).

28. Ohhata, T., Hoki, Y., Sasaki, H. & Sado, T. Crucial role of antisense transcription across the Xist promoter in Tsix-mediated Xist chromatin modification. *Development* **135,** 227–35 (2008).

29. Migeon, B. R., Chowdhury, A. K., Dunston, J. A. & McIntosh, I. Identification of TSIX, encoding an RNA antisense to human XIST, reveals differences from its murine counterpart: implications for X inactivation. *The American Journal of Human Genetics* **69,** 951–960 (2001).

30. Vallot, C., Huret, C., Lesecque, Y., Resch, A., Oudrhiri, N., Bennaceur-Griscelli, A., Duret, L. & Rougeulle, C. XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells. *Nature Genetics* **45,** 239–241 (2013).

31. Vallot, C., Patrat, C., Collier, A. J., Huret, C., Casanova, M., Ali, T. M. L., Tosolini, M., Frydman, N., Heard, E., Rugg-Gunn, P. J., *et al.* XACT noncoding RNA competes with XIST in the control of X chromosome activity during human early development. *Cell Stem Cell* **20,** 102–111 (2017).

32. Ogawa, Y. & Lee, J. T. Xite, X-Inactivation Intergenic Transcription Elements that Regulate the Probability of Choice. *Molecular Cell* **11,** 731–743 (2003).

33. Stavropoulos, N., Rowntree, R. K. & Lee, J. T. Identification of developmentally specific enhancers for Tsix in the regulation of X chromosome inactivation. *Molecular and Cellular Biology* **25,** 2757–69 (2005).

34. Galupa, R., Nora, E. P., Worsley-Hunt, R., Picard, C., Gard, C., van Bemmel, J. G., Servant, N., Zhan, Y., El Marjou, F., Johanneau, C., *et al.* A conserved noncoding locus regulates random monoallelic xist expression across a topological boundary. *Molecular Cell* **77,** 352–367 (2020).

35. Jonkers, I., Barakat, T. S., Achame, E. M., Monkhorst, K., Kenter, A., Rentmeester, E., Grosveld, F., Grootegoed, J. A. & Gribnau, J. RNF12 Is an X-Encoded Dose-Dependent Activator of X Chromosome Inactivation. *Cell* **139,** 999–1011 (2009).

36. Gontan, C., Achame, E. M., Demmers, J., Barakat, T. S., Rentmeester, E., Van Ijcken, W., Grootegoed, J. A. & Gribnau, J. RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. *Nature* **485,** 386–390 (2012).

37. Gontan, C., Mira-Bontenbal, H., Magaraki, A., Dupont, C., Barakat, T. S., Rentmeester, E., Demmers, J. & Gribnau, J. REX1 is the critical target of RNF12 in imprinted X chromosome inactivation in mice. *Nature Communications* **9,** 4752 (2018).

38. Barakat, T. S., Gunhanlar, N., Pardo, C. G., Achame, E. M., Ghazvini, M., Boers, R., Kenter, A., Rentmeester, E., Grootegoed, J. A. & Gribnau, J. RNF12 activates Xist and is essential for X chromosome inactivation. *PLoS Genetics* **7,** 1–12 (2011).

39. Shin, J., Bossenz, M., Chung, Y., Ma, H., Byron, M., Taniguchi-Ishigaki, N., Zhu, X., Jiao, B., Hall, L. L., Green, M. R., Jones, S. N., Hermans-Borgmeyer, I., Lawrence, J. B. & Bach, I. Maternal Rnf12/RLIM is required for imprinted X-chromosome inactivation in mice. *Nature* **467,** 977–981 (2010).

40. Shin, J., Wallingford, M. C., Gallant, J., Marcho, C., Jiao, B., Byron, M., Bossenz, M., Lawrence, J. B., Jones, S. N., Mager, J. & Bach, I. RLIM is dispensable for X-chromosome inactivation in the mouse embryonic epiblast. *Nature* **511,** 86–89 (2014).

41. Wang, F., McCannell, K. N., Bošković, A., Zhu, X., Shin, J., Yu, J., Gallant, J., Byron, M., Lawrence, J. B., Zhu, L. J., Jones, S. N., Rando, O. J., Fazzio, T. G. & Bach, I. Rlim -Dependent and -Independent Pathways for X Chromosome Inactivation in Female ESCs. *Cell Reports* **21,** 3691–3699 (2017).

42. Sun, S., Del Rosario, B. C., Szanto, A., Ogawa, Y., Jeon, Y. & Lee, J. T. Jpx RNA activates Xist by evicting CTCF. *Cell* **153,** 1537 (2013).

43. Barakat, T. S., Loos, F., van Staveren, S., Myronova, E., Ghazvini, M., Grootegoed, J. A. & Gribnau, J. The trans-activator RNF12 and cis-acting elements effectuate X chromosome inactivation independent of X-pairing. *Molecular Cell* **53,** 965–78 (2014).

44. Rosspopoff, O., Huret, C., Collier, A. J., Casanova, M., Rugg-Gunn, P. J., Ouimette, J.-F. & Rougeulle, C. Mechanistic diversification of XIST regulatory network in mammals. *bioRxiv,* 689430 (2019).

45. Furlan, G., Gutierrez Hernandez, N., Huret, C., Galupa, R., van Bemmel, J. G., Romito, A., Heard, E., Morey, C. & Rougeulle, C. The Ftx Noncoding Locus Controls X Chromosome Inactivation Independently of Its RNA Products. *Molecular Cell* **70,** 462–472.e8 (2018).

46. Ficz, G., Hore, T. A., Santos, F., Lee, H. J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.-L., Walter, J., *et al.* FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* **13,** 351–359 (2013).

47. Habibi, E., Brinkman, A. B., Arand, J., Kroeze, L. I., Kerstens, H. H., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N. C., *et al.* Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13,** 360–369 (2013).

48. Leitch, H. G., McEwen, K. R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J. G., Smith, A., *et al.* Naive pluripotency is associated with global DNA hypomethylation. *Nature Structural & Molecular Biology* **20,** 311 (2013).

49. Kalkan, T., Olova, N., Roode, M., Mulas, C., Lee, H. J., Nett, I., Marks, H., Walker, R., Stunnenberg, H. G., Lilley, K. S., *et al.* Tracking the embryonic stem cell transition from ground state pluripotency. *Development* **144,** 1221–1234 (2017).

50. Schulz, E. G. & Heard, E. Role and control of X chromosome dosage in mammalian development. *Current Opinion in Genetics & Development* **23,** 109–115 (2013).

51. Donohoe, M. E., Silva, S. S., Pinter, S. F., Xu, N. & Lee, J. T. The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature* **460,** 128–132 (2009).

52. Navarro, P., Chureau, C., Morey, C., Rougeulle, C., Avner, P., Chambers, I. & Karwacki-Neisius, V. Molecular coupling of Xist regulation and pluripotency. *Science* **321,** 1693–1695 (2008).

53. Navarro, P., Oldfield, A., Legoupi, J., Festuccia, N., Dubois, A., Attia, M., Schoorlemmer, J., Rougeulle, C., Chambers, I. & Avner, P. Molecular coupling of Tsix regulation and pluripotency. *Nature* **468,** 457–460 (2010).

54. Silva, J., Nichols, J., Theunissen, T. W., Guo, G., van Oosten, A. L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I. & Smith, A. Nanog Is the Gateway to the Pluripotent Ground State. *Cell* **138,** 722–737 (2009).

55. Ooi, S. K., Wolf, D., Hartung, O., Agarwal, S., Daley, G. Q., Goff, S. P. & Bestor, T. H. Dynamic instability of genomic methylation patterns in pluripotent stem cells. *Epigenetics & Chromatin* **3,** 1–10 (2010).

56. Zvetkova, I., Apedaile, A., Ramsahoye, B., Mermoud, J. E., Crompton, L. A., John, R., Feil, R. & Brockdorff, N. Global hypomethylation of the genome in XX embryonic stem cells. *Nature Genetics* **37,** 1274–1279 (2005).

57. Popp, C., Dean, W., Feng, S., Cokus, S. J., Andrews, S., Pellegrini, M., Jacobsen, S. E. & Reik, W. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463,** 1101–1105 (2010).

58. Schulz, E. G., Meisig, J., Nakamura, T., Okamoto, I., Sieber, A., Picard, C., Borensztein, M., Saitou, M., Blüthgen, N. & Heard, E. The two active X chromosomes in female ESCs block exit from the pluripotent state by modulating the ESC signaling network. *Cell Stem Cell* **14,** 203–216 (2014).

59. Norris, D. P., Patel, D., Kay, G. F., Penny, G. D., Brockdorff, N., Sheardown, S. A. & Rastan, S. Evidence that random and imprinted Xist expression is controlled by preemptive methylation. *Cell* **77,** 41–51 (1994).

60. Sado, T., Tada, T. & Takagi, N. Mosaic methylation of Xist gene before chromosome inactivation in undifferentiated female mouse embryonic stem and embryonic germ cells. *Developmental Dynamics* **205,** 421–434 (1996).

61. Beard, C., Li, E. & Jaenisch, R. Loss of methylation activates Xist in somatic but not in embryonic cells. *Genes & Development* **9,** 2325–2334 (1995).

62. Panning, B. & Jaenisch, R. DNA hypomethylation can activate Xist expression and silence X-linked genes. *Genes & Development* **10,** 1991–2002 (1996).

63. Mutzel, V. & Schulz, E. G. Dosage Sensing, Threshold Responses, and Epigenetic Memory: A Systems Biology Perspective on Random X-Chromosome Inactivation. *BioEssays* **42,** 1–14 (2020).

64. Chu, C., Zhang, Q. C., da Rocha, S. T., Flynn, R. A., Bharadwaj, M., Calabrese, J. M., Magnuson, T., Heard, E. & Chang, H. Y. Systematic discovery of Xist RNA binding proteins. *Cell* **161,** 404–16 (2015).

65. McHugh, C. A., Chen, C.-K., Chow, A., Surka, C. F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A., *et al.* The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521,** 232–236 (2015).

66. Moindrot, B., Cerase, A., Coker, H., Masui, O., Grijzenhout, A., Pintacuda, G., Schermelleh, L., Nesterova, T. B. & Brockdorff, N. A Pooled shRNA Screen Identifies Rbm15, Spen, and Wtap as Factors Required for Xist RNA-Mediated Silencing. *Cell Reports* **12,** 562–572 (2015).

67. Monfort, A., Di Minin, G., Postlmayr, A., Freimann, R., Arieti, F., Thore, S. & Wutz, A. Identification of Spen as a crucial factor for Xist function through forward genetic screening in haploid embryonic stem cells. *Cell Reports* **12,** 554–561 (2015).

68. Nesterova, T. B., Wei, G., Coker, H., Pintacuda, G., Bowness, J. S., Zhang, T., Almeida, M., Bloechl, B., Moindrot, B., Carter, E. J., Rodrigo, I. A., Pan, Q., Bi, Y., Song, C.-X. & Brockdorff, N. Systematic Allelic Analysis Defines the Interplay of Key Pathways in X Chromosome Inactivation. *Nature Communications* **10,** 1–15 (2019).

69. Coker, H., Wei, G., Moindrot, B., Mohammed, S., Nesterova, T. & Brockdorff, N. The role of the Xist 5' m6A region and RBM15 in X chromosome inactivation. *Wellcome Open Research* **5,** 31 (2020).

70. De Andrade E Sousa, L. B., Jonkers, I., Syx, L., Dunkel, I., Chaumeil, J., Picard, C., Foret, B., Chen, C. J., Lis, J. T., Heard, E., Schulz, E. G. & Marsico, A. Kinetics of Xist-induced gene silencing can be predicted from combinations of epigenetic and genomic features. *Genome Research* **29,** 1087–1099 (2019).

71. Patil, D. P., Chen, C. K., Pickering, B. F., Chow, A., Jackson, C., Guttman, M. & Jaffrey, S. R. M6 A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537,** 369–373 (2016).

72. Chen, C.-K., Blanco, M., Jackson, C., Aznauryan, E., Ollikainen, N., Surka, C., Chow, A., Cerase, A., McDonel, P. & Guttman, M. Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science* **354,** 468–472 (2016).

73. Hoki, Y., Kimura, N., Kanbayashi, M., Amakawa, Y., Ohhata, T., Sasaki, H. & Sado, T. A proximal conserved repeat in the Xist gene is essential as a genomic element for X-inactivation in mouse. *Development* **136,** 139–46 (2009).

74. Robert-Finestra, T., Tan, B. F., Mira-Bontenbal, H., Timmers, E., Gontan-Pardo, C., Merzouk, S., Giaimo, B. D., Dossin, F., van IJcken, W. F., Martens, J. W., *et al.* SPEN is Required for Xist Upregulation during Initiation of X Chromosome Inactivation. *bioRxiv* (2020).

75. Lock, L. F., Takagi, N. & Martin, G. R. Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. *Cell* **48,** 39–46 (1987).

76. Norris, D. P., Brockdorff, N. & Rastan, S. Methylation status of CpG-rich islands on active and inactive mouse X chromosomes. *Mammalian Genome* **1,** 78–83 (1991).

77. Hellman, A. & Chess, A. Gene body-specific methylation on the active X chromosome. *Science* **315,** 1141–1143 (2007).

78. Graves, J. A. M. 5-azacytidine-induced re-expression of alleles on the inactive X chromosome in a hybrid mouse cell line. *Experimental Cell Research* **141,** 99–105 (1982).

79. Sado, T., Fenner, M. H., Tan, S.-S., Tam, P., Shioda, T. & Li, E. X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. *Developmental Biology* **225,** 294–303 (2000).

80. Kalantry, S., Mills, K. C., Yee, D., Otte, A. P., Panning, B. & Magnuson, T. The Polycomb group protein Eed protects the inactive X-chromosome from differentiation-induced reactivation. *Nature Cell Biology* **8,** 195–202 (2006).

81. Pandya-Jones, A., Markaki, Y., Serizay, J., Chitiashvili, T., Leon, W. R. M., Damianov, A., Chronis, C., Papp, B., Chen, C.-K., McKee, R., *et al.* A protein assembly mediates Xist localization and gene silencing. *Nature* **587,** 145–151 (2020).

82. Okamoto, I., Arnaud, D., Le Baccon, P., Otte, A. P., Disteche, C. M., Avner, P. & Heard, E. Evidence for de novo imprinted X-chromosome inactivation independent of meiotic inactivation in mice. *Nature* **438,** 369–373 (2005).

83. Beletskii, A., Hong, Y.-K., Pehrson, J., Egholm, M. & Strauss, W. M. PNA interference mapping demonstrates functional domains in the noncoding RNA Xist. *Proceedings of the National Academy of Sciences* **98,** 9215–9220 (2001).

84. Sarma, K., Levasseur, P., Aristarkhov, A. & Lee, J. T. Locked nucleic acids (LNAs) reveal sequence requirements and kinetics of Xist RNA localization to the X chromosome. *Proceedings of the National Academy of Sciences* **107,** 22196–22201 (2010).

85. Yamada, N., Hasegawa, Y., Yue, M., Hamada, T., Nakagawa, S. & Ogawa, Y. Xist Exon 7 Contributes to the Stable Localization of Xist RNA on the Inactive X-Chromosome. *PLoS Genetics* **11,** 1–21 (2015).

86. Sunwoo, H., Colognori, D., Froberg, J. E., Jeon, Y. & Lee, J. T. Repeat E anchors Xist RNA to the inactive X chromosomal compartment through CDKN1A-interacting protein (CIZ1). *Proceedings of the National Academy of Sciences* **114,** 10654–10659 (Oct. 2017).

87. Ridings-Figueroa, R., Stewart, E. R., Nesterova, T. B., Coker, H., Pintacuda, G., Godwin, J., Wilson, R., Haslam, A., Lilley, F., Ruigrok, R., *et al.* The nuclear matrix protein CIZ1 facilitates localization of Xist RNA to the inactive X-chromosome territory. *Genes & Development* **31,** 876–888 (2017).

88. Engreitz, J. M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E. S., Plath, K. & Guttman, M. The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome. *Science* **341** (2013).

89. Carrel, L. & Willard, H. F. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **434,** 400–404 (2005).

90. Yang, F., Babak, T., Shendure, J. & Disteche, C. M. Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Research* **20,** 614–622 (2010).

91. Carrel, L. & Brown, C. J. When the lyon(Ized chromosome) roars: Ongoing expression from an inactive X chromosome. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372,** 20160355 (2017).

92. Pacini, G., Dunkel, I., Mages, N., Mutzel, V., Timmermann, B., Marsico, A. & Schulz, E. G. Integrated analysis of Xist upregulation and gene silencing at the onset of random X-chromosome inactivation at high temporal and allelic resolution. *bioRxiv,* 2020.07.20.211573 (2020).

93. Berletch, J. B., Yang, F. & Disteche, C. M. Escape from X inactivation in mice and humans. *Genome Biology* **11,** 1–7 (2010).

94. Snell, D. M. & Turner, J. M. Sex chromosome effects on male–female differences in mammals. *Current Biology* **28,** R1313–R1324 (2018).

95. Inoue, A., Jiang, L., Lu, F. & Zhang, Y. Genomic imprinting of Xist by maternal H3K27me3. *Genes & Development* **31,** 1927–1932 (2017).

96. Schultz, G. A. & Heyner, S. Gene expression in pre-implantation mammalian embryos. *Mutation Research/Reviews in Genetic Toxicology* **296,** 17–31 (1992).

97. Wells, D., Bermudez, M. G., Steuerwald, N., Thornhill, A. R., Walker, D. L., Malter, H., Delhanty, J. D. & Cohen, J. Expression of genes regulating chromosome segregation, the cell cycle and apoptosis during human preimplantation development. *Human Reproduction* **20,** 1339–1348 (2005).

98. Van den Berg, I. M., Laven, J. S. E., Stevens, M., Jonkers, I., Galjaard, R. J., Gribnau, J. & Hikke van Doorninck, J. X Chromosome Inactivation Is Initiated in Human Preimplantation Embryos. *The American Journal of Human Genetics* **84,** 771–779 (2009).

99. Okamoto, I., Patrat, C., Thépot, D., Peynot, N., Fauque, P., Daniel, N., Diabangouaya, P., Wolf, J.-P., Renard, J.-P., Duranthon, V. & Heard, E. Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature* **472,** 370–4 (2011).

100. Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S. P., Codeluppi, S., Reyes, A. P., Linnarsson, S., Sandberg, R. & Lanner, F. Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* **165,** 1012–1026 (2016).

101. An, C., Feng, G., Zhang, J., Cao, S., Wang, Y., Wang, N., Lu, F., Zhou, Q. & Wang, H. Overcoming Autocrine FGF Signaling-Induced Heterogeneity in Naive Human ESCs Enables Modeling of Random X Chromosome Inactivation. *Cell Stem Cell* **27,** 482–497 (2020).

102. Hook, E. B. & Warburton, D. The distribution of chromosomal genotypes associated with Turner's syndrome: livebirth prevalence rates and evidence for diminished fetal mortality and severity in genotypes associated with structural X abnormalities or mosaicism. *Human Genetics* **64,** 24–27 (1983).

103. Jacobs, P. A. & Strong, J. A. A case of human intersexuality having a possible XXY sex-determining mechanism. *Nature* **183,** 302–303 (1959).

104. Barr, M. L. & Carr, D. H. Sex chromatin, sex chromosomes and sex anomalies. *Canadian Medical Association Journal* **83,** 979 (1960).

105. Speirs, S., Cross, J. & Kaufman, M. The pattern of X-chromosome inactivation in the embryonic and extra-embryonic tissues of post-implantation digynic triploid LT/Sv strain mouse embryos. *Genetics Research* **56,** 107–114 (1990).

106. Endo, S., Takagi, N. & Sasaki, M. The late-replicating X chromosome in digynous mouse triploid embryos. *Developmental Genetics* **3,** 165–176 (1982).

107. Monkhorst, K., Jonkers, I., Rentmeester, E., Grosveld, F. & Gribnau, J. X Inactivation Counting and Choice Is a Stochastic Process: Evidence for Involvement of an X-Linked Activator. *Cell* **132,** 410–421 (2008).

108. Monkhorst, K., de Hoon, B., Jonkers, I., Achame, E. M., Monkhorst, W., Hoogerbrugge, J., Rentmeester, E., Westerhoff, H. V., Grosveld, F., Grootegoed, J. A., *et al.* The probability to initiate X chromosome inactivation is determined by the X to autosomal ratio and X chromosome specific allelic properties. *PLoS One* **4,** e5616 (2009).

109. Lyon, M. F. Possible mechanisms of X chromosome inactivation. *Nature New Biology* **232,** 229–232 (1971).

110. Comings, D. E. The Rationale for an Ordered Arrangement of Chromatin in the Interphase Nucleus. *The American Journal of Human Genetics* **20,** 440–460 (1968).

111. Ohno, S. Evolution of sex chromosomes in mammals. *Annual Review of Genetics* **3,** 495–524 (1969).

112. Nicodemi, M. & Prisco, A. Self-assembly and DNA binding of the blocking factor in X chromosome inactivation. *PLoS Computational Biology* **3,** e210 (2007).

113. Nicodemi, M. & Prisco, A. Symmetry-breaking model for X-chromosome inactivation. *Physical Review Letters* **98,** 108104 (2007).

114. Bacher, C. P., Guggiari, M., Brors, B., Augui, S., Clerc, P., Avner, P., Eils, R. & Heard, E. Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nature Cell Biology* **8,** 293–299 (2006).

115. Xu, N., Tsai, C. L. & Lee, J. T. Transient homologous chromosome pairing marks the onset of X inactivation. *Science* **311,** 1149–1152 (2006).

116. Augui, S., Filion, G. J., Huart, S., Nora, E. P., Guggiari, M., Maresca, M., Stewart, A. F. & Heard, E. Sensing X chromosome pairs before X inactivation via a novel X-pairing region of the Xic. *Science* **318,** 1632–1636 (2007).

117. Pollex, T. & Heard, E. Nuclear positioning and pairing of X-chromosome inactivation centers are not primary determinants during initiation of random X-inactivation. *Nature Genetics* **51,** 285–295 (2019).

118. Zhang, Q., Bhattacharya, S. & Andersen, M. E. Ultrasensitive response motifs: Basic amplifiers in molecular signalling networks. *Open Biology* **3,** 130031 (2013).

119. Ferrell, J. E. & Ha, S. H. Ultrasensitivity part II: Multisite phosphorylation, stoichiometric inhibitors, and positive feedback. *Trends in Biochemical Sciences* **39,** 556–569 (2014).

120. Buchler, N. E. & Louis, M. Molecular Titration and Ultrasensitivity in Regulatory Networks. *Journal of Molecular Biology* **384,** 1106–1119 (2008).

121. Buchler, N. E. & Cross, F. R. Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Molecular Systems Biology* **5,** 1–7 (2009).

122. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403,** 339–342 (2000).

123. Ferrell, J. E. Self-perpetuating states in signal transduction: Positive feedback, double-negative feedback and bistability. *Current Opinion in Cell Biology* **14,** 140–148 (2002).

124. Zhou, J. X. & Huang, S. Understanding gene circuits at cell-fate branch points for rational cell reprogramming. *Trends in Genetics* **27,** 55–62 (2011).

125. Ptashne, M. A genetic switch: Gene control and phage lambda. *Cell Press & Blackwell Scientific Publications* (1986).

126. Schubert, R. A., Dodd, I. B., Egan, J. B. & Shearwin, K. E. Cro's role in the CI-Cro bistable switch is critical for λ's transition from lysogeny to lytic development. *Genes & Development* **21,** 2461–2472 (2007).

127. Oppenheim, A. B., Kobiler, O., Stavans, J., Court, D. L. & Adhya, S. Switches in Bacteriophage Lambda Development. *Annual Review of Genetics* **39,** 409–429 (2005).

128. Sneppen, K. Models of life: epigenetics, diversity and cycles. *Reports on Progress in Physics* **80,** 042601 (2017).

129. Grumbach, M. M., Morishima, A. & Taylor, J. H. Human sex chromosome abnormalities in relation to DNA replication and heterochromatinization. *Proceedings of the National Academy of Sciences* **49,** 581 (1963).

130. Dodd, I. B., Micheelsen, M. A., Sneppen, K. & Thon, G. Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell* **129,** 813–22 (2007).

131. Chen, Z. X. & Riggs, A. D. Maintenance and regulation of DNA methylation patterns in mammals. *Biochemistry and Cell Biology* **83,** 438–448 (2005).

132. Rose, N. R. & Klose, R. J. Understanding the relationship between DNA methylation and histone lysine methylation. *Biochimica et Biophysica Acta* **1839,** 1362–1372 (2014).

133. Yang, H., Howard, M. & Dean, C. Antagonistic roles for H3K36me3 and H3K27me3 in the cold-induced epigenetic switch at Arabidopsis FLC. *Current Biology* **24,** 1793–1797 (2014).

134. Sun, B. K., Deaton, A. M. & Lee, J. T. A Transient Heterochromatic State in Xist Preempts X Inactivation Choice without RNA Stabilization. *Molecular Cell* **21,** 617–628 (2006).

135. Gillespie, D. T. A general method for numerically simulating coupled chemical reactions. *Journal of Computational Physics* **22,** 403–434 (1976).

136. Gillespie, D. T. Exact Stochastic Simulation of Coupled Chemical Reactions. *The Journal of Physical Chemistry* **81,** 2340–2361 (1977).

137. Henery, C., Bard, J. B. & Kaufman, M. H. Tetraploidy in mice, embryonic cell number, and the grain of the developmental map. *Developmental Biology* **152,** 233–241 (1992).

138. Webb, S., De Vries, T. & Kaufman, M. The differential staining pattern of the X chromosome in the embryonic and extraembryonic tissues of postimplantation homozygous tetraploid mouse embryos. *Genetics Research* **59,** 205–214 (1992).

139. Pintacuda, G. & Cerase, A. X Inactivation Lessons from Differentiating Mouse Embryonic Stem Cells. *Stem Cell Reviews and Reports* **11,** 699–705 (2015).

140. Guyochin, A., Maenner, S., Chu, E. T. J., Hentati, A., Attia, M., Avner, P. & Clerc, P. Live cell imaging of the nascent inactive X chromosome during the early differentiation process of naive ES cells towards epiblast stem cells. *PLoS One* **9,** 1–20 (2014).

141. Sousa, E. J., Stuart, H. T., Bates, L. E., Ghorbani, M., Nichols, J., Dietmann, S. & Silva, J. C. Exit from Naive Pluripotency Induces a Transient X Chromosome Inactivation-like State in Males. *Cell Stem Cell* **22,** 919–928 (2018).

142. Sakata, Y., Nagao, K., Hoki, Y., Sasaki, H., Obuse, C. & Sado, T. Defects in dosage compensation impact global gene regulation in the mouse trophoblast. *Development* **144,** 2784–2797 (2017).

143. Burgoyne, P. S., Thornhill, A., Boudrean, S. K., Darling, S., Bishop, C. & Evans, E. The genetic basis of XX-XY differences present before gonadal sex differentiation in the mouse. *Philosophical Transactions of the Royal Society B: Biological Sciences* **350,** 253–261 (1995).

144. Thornhill, A. R. & Burgoyne, P. S. A paternally imprinted X chromosome retards the development of the early mouse embryo. *Development* **118,** 171–174 (1993).

145. Mittwoch, U. Blastocysts prepare for the race to be male. *Human Reproduction* **8,** 1550–1555 (1993).

146. Song, J., Janiszewski, A., De Geest, N., Vanheer, L., Talon, I., El Bakkali, M., Oh, T. & Pasque, V. X-chromosome dosage modulates multiple molecular and cellular properties of mouse pluripotent stem cells independently of global DNA methylation levels. *Stem Cell Reports* **12,** 333–350 (2019).

147. Genolet, O., Monaco, A. A., Dunkel, I., Boettcher, M. & Schulz, E. G. Identification of X-chromosomal genes that drive global X-dosage effects in mammals. *bioRxiv* (2020).

148. Aeby, E., Lee, H.-G., Lee, Y.-W., Kriz, A., Del Rosario, B. C., Oh, H. J., Boukhali, M., Haas, W. & Lee, J. T. Decapping enzyme 1A breaks X-chromosome symmetry by controlling Tsix elongation and RNA turnover. *Nature Cell Biology,* 1–14 (2020).

149. Jonkers, I., Kwak, H. & Lis, J. T. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* **3,** e02407 (2014).

150. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123,** 1025–1036 (2005).

151. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biology* **4,** e309 (2006).

152. Ward, D. F. & Murray, N. E. Convergent transcription in bacteriophage λ: Interference with gene expression. *Journal of Molecular Biology* **133,** 249–266 (1979).

153. Crampton, N., Bonass, W. A., Kirkham, J., Rivetti, C. & Thomson, N. H. Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. *Nucleic Acids Research* **34,** 5416–5425 (2006).

154. Hobson, D. J., Wei, W., Steinmetz, L. M. & Svejstrup, J. Q. RNA Polymerase II Collision Interrupts Convergent Transcription. *Molecular Cell* **48,** 365–374 (2012).

155. Mak, W., Nesterova, T. B., De Napoles, M., Appanah, R., Yamanaka, S., Otte, A. P. & Brockdorff, N. Reactivation of the Paternal X Chromosome in Early Mouse Embryos. *Science* **303,** 666–669 (2004).

156. Okamoto, I., Otte, A. P., Allis, C. D., Reinberg, D. & Heard, E. Epigenetic Dynamics of Imprinted X Inactivation During Early Mouse Development. *Science* **303,** 644–649 (2004).

157. Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **379,** 131–137 (1996).

158. Lee, J. T. Homozygous Tsix mutant mice reveal a sex-ratio distortion and revert to random X-inactivation. *Nature Genetics* **32,** 195–200 (2002).

159. Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462,** 587–594 (2009).

160. Brandman, O., Ferrell, J. E., Li, R. & Meyer, T. Systems biology: Interlinked fast and slow positive feedback loops drive reliable cell decisions. *Science* **310,** 496–498 (2005).

161. Becskei, A., Séraphin, B. & Serrano, L. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *The EMBO Journal* **20,** 2528–2535 (2001).

162. Abrieu, A., Dorée, M. & Fisher, D. The interplay between cyclin-B-Cdc2 kinase (MPF) and MAP kinase during maturation of oocytes. *Journal of Cell Science* **114,** 257–267 (2001).

163. Wedlich-Soldner, R., Wai, S. C., Schmidt, T. & Li, R. Robust cell polarity is a dynamic state established by coupling transport and GTPase signaling. *Journal of Cell Biology* **166,** 889–900 (2004).

164. Gartler, S. M. & Riggs, A. D. Mammalian x-chromosome inactivation. *Annual Review of Genetics* **17,** 155–190 (1983).

165. Mutzel, V., Okamoto, I., Dunkel, I., Saitou, M., Giorgetti, L., Heard, E. & Schulz, E. G. A symmetric toggle switch explains the onset of random X inactivation in different mammals. *Nature Structural & Molecular Biology* **26,** 350–360 (2019).

166. Shiura, H. & Abe, K. Xist/Tsix expression dynamics during mouse peri-implantation development revealed by whole-mount 3D RNA-FISH. *Scientific Reports* **9,** 1–11 (2019).

167. Schulz, E. G. X-chromosome dosage as a modulator of pluripotency, signalling and differentiation? *Philosophical Transactions of the Royal Society B: Biological Sciences* **372,** 20160366 (2017).

168. Alsing, A. K. & Sneppen, K. Differentiation of developing olfactory neurons analysed in terms of coupled epigenetic landscapes. *Nucleic Acids Research* **41,** 4755–4764 (2013).

169. Serizawa, S., Ishii, T., Nakatani, H., Tsuboi, A., Nagawa, F., Asano, M., Sudo, K., Sakagami, J., Sakano, H., Ijiri, T., *et al.* Mutually exclusive expression of odorant receptor transgenes. *Nature Neuroscience* **3,** 687–693 (2000).

170. Shykind, B. M., Rohani, S. C., O'Donnell, S., Nemes, A., Mendelsohn, M., Sun, Y., Axel, R. & Barnea, G. Gene switching and the stability of odorant receptor gene choice. *Cell* **117,** 801–815 (2004).

171. Ferrell, J. E. & Machleder, E. M. The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes. *Science* **280,** 895–898 (1998).

172. Gil, N. & Ulitsky, I. Regulation of gene expression by cis-acting long non-coding RNAs. *Nature Reviews Genetics* **21,** 102–117 (2020).

173. Lee, J., Davidow, L. S. & Warshawsky, D. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nature Genetics* **21,** 400–404 (1999).

174. Vigneau, S., Augui, S., Navarro, P., Avner, P. & Clerc, P. An essential role for the DXPas34 tandem repeat and Tsix transcription in the counting process of X chromosome inactivation. *Proceedings of the National Academy of Sciences* **103,** 7390–7395 (2006).

175. Clerc, P. & Avner, P. Role of the region 3' to Xist exon 6 in the counting process of X-chromosome inactivation. *Nature Genetics* **19,** 249–253 (1998).

176. Morey, C., Navarro, P., Debrand, E., Avner, P., Rougeulle, C. & Clerc, P. The region 3' to Xist mediates X chromosome counting and H3 Lys-4 dimethylation within the Xist gene. *The EMBO Journal* **23,** 594–604 (2004).

177. Chapman, A. G., Cotton, A. M., Kelsey, A. D. & Brown, C. J. Differentially methylated CpG island within human XIST mediates alternative P2 transcription and YY1 binding. *BMC Genetics* **15,** 1–11 (2014).

178. Makhlouf, M., Ouimette, J. F., Oldfield, A., Navarro, P., Neuillet, D. & Rougeulle, C. A prominent and conserved role for YY1 in Xist transcriptional activation. *Nature Communications* **5,** 1–12 (2014).

179. Kim, J. D., Faulk, C. & Kim, J. Retroposition and evolution of the DNA-binding motifs of YY1, YY2 and REX1. *Nucleic Acids Research* **35,** 3442–3452 (2007).

180. Brown, C. J. & Willard, H. F. The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. *Nature* **368,** 154–156 (1994).

181. Csankovszki, G., Panning, B., Bates, B., Pehrson, J. R. & Jaenisch, R. Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. *Nature Genetics* **22,** 323–324 (1999).

182. Wutz, A. & Jaenisch, R. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Molecular Cell* **5,** 695–705 (2000).

183. Duffié, R. & Bourc'his, D. Parental epigenetic asymmetry in mammals. *Current Topics in Developmental Biology* **104,** 293–328 (2013).

184. Greenberg, M. V. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology* **20,** 590–607 (2019).

185. Sado, T., Okano, M., Li, E. & Sasaki, H. De novo DNA methylation is dispensable for the initiation and propagation of X chromosome inactivation. *Development* **131,** 975–982 (2004).

186. Sigal, A., Milo, R., Cohen, A., Geva-Zatorsky, N., Klein, Y., Liron, Y., Rosenfeld, N., Danon, T., Perzov, N. & Alon, U. Variability and memory of protein levels in human cells. *Nature* **444,** 643–646 (2006).

187. Chureau, C., Prissette, M., Bourdet, A., Barbe, V., Cattolico, L., Jones, L., Eggen, A., Avner, P. & Duret, L. Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. *Genome Research* **12,** 894–908 (2002).

188. Tian, D., Sun, S. & Lee, J. T. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell* **143,** 390–403 (2010).

189. Chureau, C., Chantalat, S., Romito, A., Galvani, A., Duret, L., Avner, P. & Rougeulle, C. Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. *Human Molecular Genetics* **20,** 705–718 (2011).

190. Rayon, T., Stamataki, D., Perez-Carrasco, R., Garcia-Perez, L., Barrington, C., Melchionda, M., Exelby, K., Lazaro, J., Tybulewicz, V. L., Fisher, E. M., *et al.* Species-specific pace of development is associated with differences in protein stability. *Science* **369** (2020).

191. Matsuda, M., Yamanaka, Y., Uemura, M., Osawa, M., Saito, M. K., Nagahashi, A., Nishio, M., Guo, L., Ikegawa, S., Sakurai, S., *et al.* Recapitulating the human segmentation clock with pluripotent stem cells. *Nature* **580,** 124–129 (2020).

192. Mukherjee, B. B. & Sinha, A. K. Single-active-X hypothesis: Cytological evidence for random inactivation of X-chromosomes in a female mule complement. *Proceedings of the National Academy of Sciences* **51,** 252 (1964).

193. Wang, X., Miller, D. C., Clark, A. G. & Antczak, D. F. Random X inactivation in the mule and horse placenta. *Genome Research* **22,** 1855–1863 (2012).

194. Migeon, B. R., Kazi, E., Haisley-royster, C., Hu, J., Reeves, R., Call, L., Lawler, A., Moore, C. S., Morrison, H. & Jeppesen, P. Human X inactivation center induces random X chromosome inactivation in male transgenic mice. *Genomics* **59,** 113–121 (1999).

195. Heard, E., Mongelard, F., Arnaud, D., Chureau, C., Vourc'h, C. & Avner, P. Human XIST yeast artificial chromosome transgenes show partial X inactivation center function in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences* **96,** 6841–6846 (1999).

196. Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C., Suzuki, M., Kawai, J., *et al.* Antisense transcription in the mammalian transcriptome. *Science* **309,** 1564–1566 (2005).

197. Yelin, R., Dahary, D., Sorek, R., Levanon, E. Y., Goldstein, O., Shoshan, A., Diber, A., Biton, S., Tamir, Y., Khosravi, R., *et al.* Widespread occurrence of antisense transcription in the human genome. *Nature Biotechnology* **21,** 379–386 (2003).

198. Misra, S., Crosby, M. A., Mungall, C. J., Matthews, B. B., Campbell, K. S., Hradecky, P., Huang, Y., Kaminker, J. S., Millburn, G. H., Prochnik, S. E., *et al.* Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biology* **3,** 1–22 (2002).

199. Wang, X. J., Gaasterland, T. & Chua, N. H. Genome-wide prediction and identification of cis-natural antisense transcripts in Arabidopsis thaliana. *Genome Biology* **6,** 1–11 (2005).

200. David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C. J., Bofkin, L., Jones, T., Davis, R. W. & Steinmetz, L. M. A high-resolution map of transcription in the yeast genome. *Proceedings of the National Academy of Sciences* **103,** 5320–5325 (2006).

201. Dornenburg, J. E., DeVita, A. M., Palumbo, M. J. & Wade, J. T. Widespread antisense transcription in Escherichia coli. *mBio* **1,** 1–4 (2010).

202. Magistri, M., Faghihi, M. A., St, G. & Iii, L. Regulation of chromatin structure by long noncoding RNAs : focus on natural antisense transcripts. *Trends in Genetics* **28,** 389–396 (2012).

203. Callen, B. P., Shearwin, K. E. & Egan, J. B. Transcriptional Interference between Convergent Promoters Caused by Elongation over the Promoter. *Molecular Cell* **14,** 647–656 (2004).

204. Engström, P. G., Suzuki, H., Ninomiya, N., Akalin, A., Sessa, L., Lavorgna, G., Brozzi, A., Luzi, L., Tan, S. L., Yang, L., *et al.* Complex loci in human and mouse genomes. *PLoS Genetics* **2,** e47 (2006).

205. Loos, F., Loda, A., Van Wijk, L., Grootegoed, J. A. & Gribnau, J. Chromatin-mediated reversible silencing of sense-antisense gene pairs in embryonic stem cells is consolidated upon differentiation. *Molecular and Cellular Biology* **35,** 2436–2447 (2015).

206. Mercer, T. R., Gerhardt, D. J., Dinger, M. E., Crawford, J., Trapnell, C., Jeddeloh, J. A., Mattick, J. S. & Rinn, J. L. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nature Biotechnology* **30,** 99–104 (2012).

207. Werner, A., Carlile, M. & Swan, D. What do natural antisense transcripts regulate? *RNA Biology* **6,** 43–48 (2009).

208. Chen, J., Sun, M., Hurst, L. D., Carmichael, G. G. & Rowley, J. D. Genome-wide analysis of coordinate expression and evolution of human cis-encoded sense-antisense transcripts. *Trends in Genetics* **21,** 326–329 (2005).

209. Yassour, M., Pfiffner, J., Levin, J. Z., Adiconis, X., Gnirke, A., Nusbaum, C., Thompson, D. A., Friedman, N. & Regev, A. Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. *Genome Biology* **11,** 1–14 (2010).

210. Mayer, A., Landry, H. M. & Churchman, L. S. Pause & go : from the discovery of RNA polymerase pausing to its functional implications. *Current Opinion in Cell Biology* **46,** 72–80 (2017).

211. Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews Molecular Cell Biology* **19,** 621–637 (2018).

212. Chubb, J. R., Trcek, T., Shenoy, S. M. & Singer, R. H. Transcriptional pulsing of a developmental gene. *Current Biology* **16,** 1018–1025 (2006).

213. Tantale, K., Mueller, F., Kozulic-Pirher, A., Lesne, A., Victor, J.-M., Robert, M.-C., Capozi, S., Chouaib, R., Bäcker, V., Mateos-Langerak, J., *et al.* A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. *Nature Communications* **7,** 1–14 (2016).

214. Bartman, C. R., Hsu, S. C., Hsiung, C. C.-S., Raj, A. & Blobel, G. A. Enhancer regulation of transcriptional bursting parameters revealed by forced chromatin looping. *Molecular Cell* **62,** 237–247 (2016).

215. Fukaya, T., Lim, B. & Levine, M. Enhancer control of transcriptional bursting. *Cell* **166,** 358–368 (2016).

216. Larsson, A. J., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O. R., Reinius, B., Segerstolpe, Å., Rivera, C. M., Ren, B. & Sandberg, R. Genomic encoding of transcriptional burst kinetics. *Nature* **565,** 251–254 (2019).

217. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature Reviews Genetics* **13,** 720–731 (2012).

218. Li, J. & Gilmour, D. S. Promoter proximal pausing and the control of gene expression. *Current Opinion in Genetics & Development* **21,** 231–235 (2011).

219. Krebs, A. R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L. & Schübeler, D. Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Molecular Cell* **67,** 411–422.e4 (2017).

220. Ehrensberger, A. H., Kelly, G. P. & Svejstrup, J. Q. Mechanistic interpretation of promoter-proximal peaks and RNAPII density maps. *Cell* **154,** 713–715 (2013).

221. Shearwin, K. E., Callen, B. P. & Egan, J. B. Transcriptional interference - A crash course. *TRENDS in Genetics* **21,** 339–345 (2005).

222. Prescott, E. M. & Proudfoot, N. J. Transcriptional collision between convergent genes in budding yeast. *Proceedings of the National Academy of Sciences* **99,** 8796–8801 (2002).

223. Sneppen, K., Dodd, I. B., Shearwin, K. E., Palmer, A. C., Schubert, R. A., Callen, B. P. & Egan, J. B. A mathematical model for transcriptional interference by RNA polymerase traffic in Escherichia coli. *Journal of Molecular Biology* **346,** 399–409 (2005).

224. Osato, N., Suzuki, Y., Ikeo, K. & Gojobori, T. Transcriptional interferences in cis natural antisense transcripts of humans and mice. *Genetics* **176,** 1299–1306 (2007).

225. Adhya, S. & Gottesman, M. Promoter occlusion: Transcription through a promoter may inhibit its activity. *Cell* **29,** 939–944 (1982).

226. Greger, I. H. & Proudfoot, N. J. Poly(A) signals control both transcriptional termination and initiation between the tandem GAL10 and GAL7 genes of Saccharomyces cerevisiae. *The EMBO Journal* **17,** 4771–4779 (1998).

227. Wang, P., Yang, J., Ishihama, A. & Pittard, A. Demonstration that the TyrR Protein and RNA Polymerase Complex Formed at the Divergent P3 Promoter Inhibits Binding of RNA Polymerase to the Major Promoter, P1, of the aroP Gene of Escherichia coli. *Journal of Bacteriology* **180,** 5466–5472 (1998).

228. Palmer, A. C., Barry Egan, J. & Shearwin, K. E. Transcriptional interference by RNA polymerase pausing and dislodgement of transcription factors. *Transcription* **2,** 9–14 (2011).

229. Isoda, T., Moore, A. J., He, Z., Chandra, V., Aida, M., Denholtz, M., Piet van Hamburg, J., Fisch, K. M., Chang, A. N., Fahl, S. P., Wiest, D. L. & Murre, C. Non-coding Transcription Instructs Chromatin Folding and Compartmentalization to Dictate Enhancer-Promoter Communication and T Cell Fate. *Cell* **171,** 103–119.e18 (2017).

230. Heinz, S., Texari, L., Hayes, M. G., Urbanowski, M., Chang, M. W., Givarkes, N., Rialdi, A., White, K. M., Albrecht, R. A., Pache, L., Marazzi, I., García-Sastre, A., Shaw, M. L. & Benner, C. Transcription Elongation Can Affect Genome 3D Structure. *Cell* **174,** 1522–1536.e22 (2018).

231. Imamura, T., Yamamoto, S., Ohgane, J., Hattori, N., Tanaka, S. & Shiota, K. Non-coding RNA directed DNA demethylation of Sphk1 CpG island. *Biochemical and Biophysical Research Communications* **322,** 593–600 (2004).

232. Sims, R. J., Belotserkovskaya, R. & Reinberg, D. Elongation by RNA polymerase II: The short and long of it. *Genes & Development* **18,** 2437–2468 (2004).

233. Baubec, T., Colombo, D. F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A. R., Akalin, A. & Schübeler, D. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520,** 243–247 (2015).

234. Morselli, M., Pastor, W. A., Montanini, B., Nee, K., Ferrari, R., Fu, K., Bonora, G., Rubbi, L., Clark, A. T., Ottonello, S., *et al.* In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. *eLife* **4,** e06205 (2015).

235. Wagner, E. J. & Carpenter, P. B. Understanding the language of Lys36 methylation at histone H3. *Nature Reviews Molecular Cell Biology* **13,** 115–126 (2012).

236. Teissandier, A. & Bourc'his, D. Gene body DNA methylation conspires with H3K36me3 to preclude aberrant transcription. *The EMBO Journal* **36,** 1471–1473 (2017).

237. Tufarelli, C., Sloane Stanley, J. A., Garrick, D., Sharpe, J. A., Ayyub, H., Wood, W. G. & Higgs, D. R. Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nature Genetics* **34,** 157–165 (2003).

238. Faghihi, M. A. & Wahlestedt, C. Regulatory roles of natural antisense transcripts. *Nature Reviews Molecular Cell Biology* **10,** 637–643 (2009).

239. Pandey, R. R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-DiNardo, D. & Kanduri, C. Kcnq1ot1 Antisense Noncoding RNA Mediates Lineage-Specific Transcriptional Silencing through Chromatin-Level Regulation. *Molecular Cell* **32,** 232–246 (2008).

240. Faghihi, M. A., Modarresi, F., Khalil, A. M., Wood, D. E., Sahagan, B. G., Morgan, T. E., Finch, C. E., St. Laurent, G., Kenny, P. J. & Wahlestedt, C. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of β-secretase. *Nature Medicine* **14,** 723–730 (2008).

241. Delaval, K. & Feil, R. Epigenetic regulation of mammalian genomic imprinting. *Current Opinion in Genetics & Development* **14,** 188–195 (2004).

242. Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A. P. & Cui, H. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* **451,** 202–206 (2008).

243. Morris, K. V., Santoso, S., Turner, A. M., Pastori, C. & Hawkins, P. G. Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. *PLoS Genetics* **4,** 3–11 (2008).

244. Zhao, J., Ohsumi, T. K., Kung, J. T., Ogawa, Y., Grau, D. J., Sarma, K., Song, J. J., Kingston, R. E., Borowsky, M. & Lee, J. T. Genome-wide Identification of Polycomb-Associated RNAs by RIP-seq. *Molecular Cell* **40,** 939–953 (2010).

245. Bernstein, E. & Allis, C. D. RNA meets chromatin. *Genes & Development* **19,** 1635–1655 (2005).

246. Latos, P. A., Pauler, F. M., Koerner, M. V., Şenergin, H. B., Hudson, Q. J., Stocsits, R. R., Allhoff, W., Stricker, S. H., Klement, R. M., Warczok, K. E., *et al.* Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* **338,** 1469–1472 (2012).

247. Kotake, Y., Nakagawa, T., Kitagawa, K., Suzuki, S., Liu, N., Kitagawa, M. & Xiong, Y. Long non-coding RNA ANRIL is required for the PRC2 recruitment to and silencing of p15 INK4B tumor suppressor gene. *Oncogene* **30,** 1956–1962 (2011).

248. Huang, M. D., Chen, W. M., Qi, F. Z., Xia, R., Sun, M., Xu, T. P., Yin, L., Zhang, E. B., De, W. & Shu, Y. Q. Long non-coding RNA ANRIL is upregulated in hepatocellular carcinoma and regulates cell proliferation by epigenetic silencing of KLF2. *Journal of Hematology & Oncology* **8,** 1–14 (2015).

249. Swiezewski, S., Liu, F., Magusin, A. & Dean, C. Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nature* **462,** 799–802 (2009).

250. Ietswaart, R., Wu, Z. & Dean, C. Flowering time control: Another window to the connection between antisense RNA and chromatin. *Trends in Genetics* **28,** 445–453 (2012).

251. Papenfort, K. & Vogel, J. Multiple target regulation by small noncoding RNAs rewires gene expression at the post-transcriptional level. *Research in Microbiology* **160,** 278–287 (2009).

252. Waters, L. S. & Storz, G. Regulatory RNAs in Bacteria. *Cell* **136,** 615–628 (2009).

253. Faghihi, M. A., Zhang, M., Huang, J., Modarresi, F., Van der Brug, M. P., Nalls, M. A., Cookson, M. R., St-Laurent, G. & Wahlestedt, C. Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biology* **11,** 1–13 (2010).

254. Lasa, I., Toledo-Arana, A., Dobin, A., Villanueva, M., De Los Mozos, I. R., Vergara-Irigaray, M., Segura, V., Fagegaltier, D., Penadés, J. R., Valle, J., Solano, C. & Gingeras, T. R. Genome-wide antisense transcription drives mRNA processing in bacteria. *Proceedings of the National Academy of Sciences* **108,** 20172–20177 (2011).

255. Naville, M. & Gautheret, D. Transcription attenuation in bacteria: Theme and variations. *Briefings in Functional Genomics* **8,** 482–492 (2009).

256. Stork, M., Di Lorenzo, M., Welch, T. J. & Crosa, J. H. Transcription termination within the iron transport-biosynthesis operon of Vibrio anguillarum requires an antisense RNA. *Journal of Bacteriology* **189,** 3479–3488 (2007).

257. Johnson, C. M., Manias, D. A., Haemig, H. A., Shokeen, S., Weaver, K. E., Henkin, T. M. & Dunny, G. M. Direct evidence for control of the pheromone-inducible prgQ operon of Enterococcus faecalis plasmid pCF10 by a countertranscript-driven attenuation mechanism. *Journal of Bacteriology* **192,** 1634–1642 (2010).

258. Pugh, B. F. & Venters, B. J. Genomic organization of human transcription initiation complexes. *PloS One* **11,** e0149339 (2016).

259. Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology* **16,** 167–177 (2015).

260. Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes & Development* **33,** 960–982 (2019).

261. Harper, C. V., Finkenstädt, B., Woodcock, D. J., Friedrichsen, S., Semprini, S., Ashall, L., Spiller, D. G., Mullins, J. J., Rand, D. A., Davis, J. R., *et al.* Dynamic analysis of stochastic transcription cycles. *PLoS Biology* **9,** e1000607 (2011).

262. Suter, D. M., Molina, N., Gatfield, D., Schneider, K., Schibler, U. & Naef, F. Mammalian genes are transcribed with widely different bursting kinetics. *Science* **332,** 472–474 (2011).

263. Coulon, A., Chow, C. C., Singer, R. H. & Larson, D. R. Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nature Reviews Genetics* **14,** 572–584 (2013).

264. Zoller, B., Nicolas, D., Molina, N. & Naef, F. Structure of silent transcription intervals and noise characteristics of mammalian genes. *Molecular Systems Biology* **11,** 823 (2015).

265. Zhang, J., Chen, L. & Zhou, T. Analytical distribution and tunability of noise in a model of promoter progress. *Biophysical Journal* **102,** 1247–1257 (2012).

266. Bartman, C. R., Hamagami, N., Keller, C. A., Giardine, B., Hardison, R. C., Blobel, G. A. & Raj, A. Transcriptional Burst Initiation and Polymerase Pause Release Are Key Control Points of Transcriptional Regulation. *Molecular Cell* **73,** 519–532.e4 (2019).

267. Henriques, T., Gilchrist, D. A., Nechaev, S., Bern, M., Muse, G. W., Burkholder, A., Fargo, D. C. & Adelman, K. Stable pausing by rna polymerase II provides an opportunity to target and integrate regulatory signals. *Molecular Cell* **52,** 517–528 (2013).

268. Alt, F. W., Oltz, E. M., Young, F., Gorman, J., Taccioli, G. & Chen, J. VDJ recombination. *Immunology Today* **13,** 306–314 (1992).

269. Chang, D.-E., Leung, S., Atkinson, M. R., Reifler, A., Forger, D. & Ninfa, A. J. Building biological memory by linking positive feedback loops. *Proceedings of the National Academy of Sciences* **107,** 175–180 (2010).

270. Xiong, W. & Ferrell, J. E. A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature* **426,** 460–465 (2003).

271. Ge, X., Wu, Q., Jung, Y. C., Chen, J. & Wang, S. M. A large quantity of novel human antisense transcripts detected by LongSAGE. *Bioinformatics* **22,** 2475–2479 (2006).

272. Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G., *et al.* Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308,** 1149–1154 (2005).

273. Brown, T., Howe, F. S., Murray, S. C., Wouters, M., Lorenz, P., Seward, E., Rata, S., Angel, A. & Mellor, J. Antisense transcription-dependent chromatin signature modulates sense transcript dynamics. *Molecular Systems Biology* **14,** 1–21 (2018).

274. Xu, Z., Wei, W., Gagneur, J., Clauder-Münster, S., Smolik, M., Huber, W. & Steinmetz, L. M. Antisense expression increases gene expression variability and locus interdependency. *Molecular Systems Biology* **7,** 468 (Feb. 2011).

275. Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J. A. & Churchman, L. S. Native Elongating Transcript Sequencing Reveals Human Transcriptional Activity at Nucleotide Resolution. *Cell* **161,** 541–554 (2015).

276. Navarro Gonzalez, J., Zweig, A. S., Speir, M. L., Schmelter, D., Rosenbloom, K. R., Raney, B. J., Powell, C. C., Nassar, L. R., Maulding, N. D., Lee, C. M., *et al.* The UCSC Genome Browser database: 2021 update. *Nucleic Acids Research* **49,** D1046–D1057 (2020).

277. Wery, M., Gautier, C., Descrimes, M., Yoda, M., Vennin-Rendos, H., Migeot, V., Gautheret, D., Hermand, D. & Morillon, A. Native elongating transcript sequencing reveals global anti-correlation between sense and antisense nascent transcription in fission yeast. *RNA* **24,** 196–208 (2018).

278. Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J. & Cramer, P. TT-seq maps the human transient transcriptome. *Science* **352,** 1225–1228 (2016).

279. Churchman, L. S. & Weissman, J. S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469,** 368–373 (2011).

280. Nojima, T., Gomes, T., Grosso, A. R. F., Kimura, H., Dye, M. J., Dhir, S., Carmo-Fonseca, M. & Proudfoot, N. J. Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* **161,** 526–540 (2015).

281. Gressel, S., Schwalb, B., Decker, T. M., Qin, W., Leonhardt, H., Eick, D. & Cramer, P. CDK9-dependent RNA polymerase II pausing controls transcription initiation. *eLife* **6,** 1–24 (2017).

282. Gill, J. K., Maffioletti, A., Garcia-Molinero, V., Stutz, F. & Soudet, J. Fine chromatin-driven mechanism of transcription interference by antisense noncoding transcription. *Cell Reports* **31,** 107612 (2020).

283. Murray, S. C., Haenni, S., Howe, F. S., Fischl, H., Chocian, K., Nair, A. & Mellor, J. Sense and antisense transcription are associated with distinct chromatin architectures across genes. *Nucleic Acids Research* **43,** 7823–7837 (2015).

284. Nevers, A., Doyen, A., Malabat, C., Néron, B., Kergrohen, T., Jacquier, A. & Badis, G. Antisense transcriptional interference mediates condition-specific gene repression in budding yeast. *Nucleic Acids Research* **46,** 6009–6025 (2018).

285. Yap, K. L., Li, S., Muñoz-Cabello, A. M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M. J. & Zhou, M. M. Molecular Interplay of the Noncoding RNA ANRIL and Methylated Histone H3 Lysine 27 by Polycomb CBX7 in Transcriptional Silencing of INK4a. *Molecular Cell* **38,** 662–674 (2010).

286. Hongay, C. F., Grisafi, P. L., Galitski, T. & Fink, G. R. Antisense Transcription Controls Cell Fate in Saccharomyces cerevisiae. *Cell* **127,** 735–745 (2006).

287. Hisanaga, T., Okahashi, K., Yamaoka, S., Kajiwara, T., Nishihama, R., Shimamura, M., Yamato, K. T., Bowman, J. L., Kohchi, T. & Nakajima, K. A cis-acting bidirectional transcription switch controls sexual dimorphism in the liverwort. *The EMBO Journal* **38,** e100240 (2019).

288. Zhu, J. & Paul, W. E. CD4 T cells : fates , functions , and faults ASH 50th anniversary review CD4 T cells : fates , functions , and faults. *Immunobiology* **112,** 1557–1569 (2009).

289. Mariani, L., Schulz, E. G., Lexberg, M. H., Helmstetter, C., Radbruch, A., Löhning, M. & Höfer, T. Short-term memory in gene induction reveals the regulatory principle behind stochastic IL-4 expression. *Molecular Systems Biology* **6,** 1–13 (2010).

290. Light, W. H., Freaney, J., Sood, V., Thompson, A., D'Urso, A., Horvath, C. M. & Brickner, J. H. A conserved role for human Nup98 in altering chromatin structure and promoting epigenetic transcriptional memory. *PLoS Biology* **11,** e1001524 (2013).

291. D'Urso, A. & Brickner, J. H. Epigenetic transcriptional memory. *Current Genetics* **63,** 435–439 (2017).

292. Bothe, M., Buschow, R. & Meijsing, S. H. Glucocorticoid receptor activation induces gene-specific transcriptional memory and universally reversible changes in chromatin accessibility. *bioRxiv* (2021).

293. Lämke, J. & Bäurle, I. Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. *Genome Biology* **18,** 1–11 (2017).

294. Bumgarner, S. L., Dowell, R. D., Grisafi, P., Gifford, D. K. & Fink, G. R. Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast. *Proceedings of the National Academy of Sciences* **106,** 18321–18326 (2009).

295. Weiss, R., Basu, S., Hooshangi, S., Kalmbach, A., Karig, D., Mehreja, R. & Netravali, I. Genetic circuit building blocks for cellular computation, communications, and signal processing. *Natural Computing* **2,** 47–84 (2003).

296. Kramer, B. P., Viretta, A. U., Baba, M. D.-E., Aubel, D., Weber, W. & Fussenegger, M. An engineered epigenetic transgene switch in mammalian cells. *Nature Biotechnology* **22,** 867–870 (2004).

297. Pimmett, V. L., Dejean, M., Fernandez, C., Trullo, A., Bertrand, E., Radulescu, O. & Lagha, M. Quantitative imaging of transcription in living Drosophila embryos reveals the impact of core promoter motifs on promoter state dynamics. *bioRxiv* (2021).

298. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences* **99,** 12795–12800 (2002).

299. Fonseca, J. P., Bonny, A. R., Kumar, G. R., Ng, A. H., Town, J., Wu, Q. C., Aslankoohi, E., Chen, S. Y., Dods, G., Harrigan, P., *et al.* A toolkit for rapid modular construction of biological circuits in mammalian cells. *ACS Synthetic Biology* **8,** 2593–2606 (2019).

300. Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S. & Mitchell, S. E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One* **6,** e19379 (2011).

301. Sambrook, J., Fritsch, E. F., Maniatis, T., *et al.* Molecular cloning: a laboratory manual. *Cold Spring Harbor Laboratory Press* (1989).

302. Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X., Shalem, O., Cradick, T. J., Marraffini, L. A., Bao, G. & Zhang, F. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology* **31,** 827–832 (2013).

303. Chaumeil, J., Augui, S., Chow, J. C. & Heard, E. in *The nucleus* 297–308 (Springer, 2008).

304. Giorgetti, L., Galupa, R., Nora, E. P., Piolot, T., Lam, F., Dekker, J., Tiana, G. & Heard, E. Predictive Polymer Modeling Reveals Coupled Fluctuations in Chromosome Conformation and Transcription. *Cell* **157,** 950–963 (2014).

305. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T. R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).

306. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).

307. Chow, J. C., Ciaudo, C., Fazzari, M. J., Mise, N., Servant, N., Glass, J. L., Attreed, M., Avner, P., Wutz, A., Barillot, E., Greally, J. M., Voinnet, O. & Heard, E. LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell* **141,** 956–969 (2010).

308. Borensztein, M., Syx, L., Ancelin, K., Diabangouaya, P., Picard, C., Liu, T., Liang, J.-B., Vassilev, I., Galupa, R., Servant, N., Barillot, E., Surani, A., Chen, C.-J. & Heard, E. Xist-dependent imprinted X inactivation and the early developmental consequences of its failure. *Nature Structural & Molecular Biology* **24,** 226–233 (2017).

309. Feil, R., Wagner, J., Metzger, D. & Chambon, P. Regulation of Cre Recombinase Activity by Mutated Estrogen Receptor Ligand-Binding Domains. *Biochemical and Biophysical Research Communications* **237,** 752–757 (1997).

310. Banaszynski, L. A., Chen, L.-c., Maynard-smith, L. A., Ooi, A. G. L. & Wandless, T. J. A Rapid , Reversible , and Tunable Method to Regulate Protein Function in Living Cells Using Synthetic Small Molecules. *Cell* **126,** 995–1004 (2006).

311. Akhtar, W., Jong, J. D., Pindyurin, A. V., Pagie, L., Meuleman, W., Ridder, J. D., Berns, A., Wessels, L. F. A., Lohuizen, M. V. & Steensel, B. V. Chromatin Position Effects Assayed by Thousands of Reporters Integrated in Parallel. *Cell* **154,** 914–927 (2013).

10 Acknowledgements

First and foremost, I would like to thank my mentor Dr. Edda Schulz. Thank you, Edda, for letting me work on this fascinating topic, for all the support you have given me, and for all the enthusiasm you put into the project. Thank you for the many stimulating discussions we have had, for pushing me to grow scientifically, and for giving me the chance to attend numerous international conferences.

I thank Prof. Dr. Hanspeter Herzel for serving as my second supervisor, for his many ideas and useful suggestions during my committee meetings, and for keeping track of my progress.

I thank Dr. Luca Giorgetti for serving as a committee member and reviewer, for his readiness to share data, reagents and ideas, for his patience in replying to countless emails and his willingness to discuss whenever needed.

A big thanks to all past and present members of the Schulz group for all the discussions, and all the small and big assistance. Thanks for so often saying yes to the countless *are you here on the weekend?* and *could you change my medium?*'s! Thank you, Liat and Oriana, for all your wet lab help and guidance, for your willingness to discuss and give input, for the many coffees, pizzas and laughs that we have shared. I thank Ilona, for all her help with experiments, for her patience and support, and for always being willing to share her experience. Thank you Guido, for all the beers that we have shared, for countless “Nummer 33” 's, and for your patience in answering and discussing all my questions. Zeba, thank you for your constant willingness to help, and for all the fun we have had in the lab or on retreats and conferences. Rutger, for making the office a more fun (and a greener) place, and for sharing his knowledge whenever asked. Till, for the many fun XCI discussions and for actually reading this complete thesis.

Science is a collaborative (ad)venture, and so this thesis would not have been possible without the help, work and data of others: I thank Edda Schulz, Ilona Dunkel, Luca Giorgetti, Ikuhiro Okamoto and Benedikt Boesen for their contributions to this work.

I also thank Alex Tuck, Alain Bonny and Anton Wutz for sharing plasmids and cell lines.

I am very grateful for the support, both financial and intellectual, of the graduate school "Computational Systems Biology" (DFG). I thank my colleagues from the CSB, many of whom have become my friends over the years, for all the scientific and non-scientific time we spent together.

"The first draft of anything is shit." Ernest Hemingway
And so I want to thank the many people who took the trouble to read versions of the manuscript. Thank you, Edda, Rupert, Sophie, Philipp, Marjan, Kirsten, Till and Simon!

Last but not least, I want to thank my family: My mom, Helga, and my dad, Rupert, who took me to the lab at the age of three and got me fascinated in biology, my sister Sophie, and you, Simon, for always being there. Thank you for your constant support, for enduring my occassional frustrations, and for giving me balance and keeping things in perspective!

11 List of publications

- Mutzel, V., Okamoto, I., Dunkel, I., Saitou, M., Giorgetti, L., Heard, E. & Schulz, E. G. A symmetric toggle switch explains the onset of random X inactivation in different mammals. *Nature Structural & Molecular Biology* **26**, 350–360 (2019).
- Mutzel, V. & Schulz, E. G. Dosage Sensing, Threshold Responses, and Epigenetic Memory: A Systems Biology Perspective on Random X-Chromosome Inactivation. *BioEssays* **42**, 1–14 (2020).
- Pacini, G., Dunkel, I., Mages, N., Mutzel, V., Timmermann, B., Marsico, A., & Schulz, E. G. Integrated analysis of Xist upregulation and X-chromosome inactivation with single-cell and single-allele resolution. *Nature Communications*, **12(1)**, 1-17 (2021).
- Gjaltema, R. A., Schwämmle, T., Kautz, P., Robson, M., Schoepflin, R., Lustig, L. R., Brandenburg, L., Dunkel, I., Vechiatto, C., Ntini, E., Mutzel, V., Schmiedel, V., Marsico, A., Mundlos, S. & Schulz, E. G. Distal and proximal cis-regulatory elements sense X-chromosomal dosage and developmental state at the Xist locus. *bioRxiv*, (2021).

12 Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Date	Verena Mutzel